이학박사학위논문

Improving genome-wide specificity and targeting-scope of CRISPR RNA-guided base editing

2019년 2월

서울대학교 대학원

화학부 생화학 전공

임 가 영

Abstract

Improving genome-wide specificity and targeting-scope of CRISPR RNA-guided base editing

Kayeong Lim

Department of Chemistry

The Graduate School

Seoul National University

Over the past few years, developments of applications in the CRISPR-Cas9 system have increased explosively, not only for efficient genome engineering but also for recruiting variety range of functional domains at a target locus. In the presence of an exogenous DNA template with homologous sequence, point mutations can be introduced at Cas9 nuclease-mediated DSBs by homology-directed repair (HDR), one of endogenous cellular repair mechanism. However, the efficiency of HDR is modest since competitive non-homologous end joining (NHEJ) pathway is predominant. Also, it has large variations on the cell type.

Recently developed base-editing systems present a useful orthogonal strategy for manipulating nucleotide substitutions. CRISPR RNA-guided base editors,

fusion proteins consist of a catalytically defective *Sterptococcus pyogenes* Cas9 and deaminases, convert single-nucleotide in the target DNA. Base editing systems are divided into two categories: cytosine base editors (CBEs) that convert C to T and adenosine base editors (ABEs) that convert A to G. Unlike programmable nucleases which produce small indels at a target site, base editing systems have the advantage that can induce base conversions, not relying on the DSB repair pathway. Despite broad interest in base editing, genome-wide target specificities of CRISPR RNA-guided base editors remain unknown. Several methods for off-target detection are established; however, existing methods are based on capturing of DSBs, and they are not suitable for programmable deaminases.

In this thesis, I will describe a new method for assessing genome-wide specificities of CRISPR RNA-guided base editing by modifying Digenome-seq. I validate off-targets of CBEs and confirm that CBEs are highly specific compared to CRISPR nucleases. To reduce the off-target effects, I use modified sgRNAs and observe that both extended and truncated sgRNAs improve the specificities of CRISPR RNA-guided base editors. Furthermore, extended sgRNAs show base editing in additional nucleotides at positions out of canonical base editing windows, demonstrating the expansion of targeting-scope of CRISPR RNA-guided deaminases.

Keywords: CRISPR-Cas9, Base Editor, Base-editing, off-target, genome-wide specificity, targeting-scope, extended sgRNA

Student Number: 2014-21246

Table of Contents

List of Figures

List of Tables

Ⅰ. Introduction

Programmable nucleases mediated genome engineering has been widely used to understand the function of a gene or a specific mutation in sequence (Kim and Kim, 2014). Programmable nucleases, such as zinc finger nucleases (ZFNs) (Bibikova et al., 2003; Doyon et al., 2008; Maeder et al., 2008; Urnov et al., 2005), transcriptional activator-like effector nucleases (TALENs) (Boch et al., 2009; Cermak et al., 2011; Kim et al., 2013; Miller et al., 2011; Moscou and Bogdanove, 2009) and clustered regularly interspaced short palindromic repeat (CRISPR) nucleases (Cho et al., 2013; Cong et al., 2013; Hwang et al., 2013; Jiang et al., 2013; Jinek et al., 2012; Mali et al., 2013b), create double-strand breaks (DSBs) at a specific locus in the genome. Two major pathways for repair of cellular DSBs are non-homologous end joining (NHEJ) and homologous recombination (HR) (Chapman et al., 2012; Liang et al., 1998). The predominant NHEJ pathway is an error-prone process that eliminates DSBs by direct ligation (Shrivastav et al., 2008) and mostly generates small insertions or deletions (indels) leading to frameshift mutations. In the presence of a homologous DNA template, DSB sites can be replaced through the HR using homologous DNA sequence leading to gene correction or gene insertion.

ZFNs and TALENs have a common feature that they are recombinant proteins made by artificially linking a sequence-specific DNA-binding protein, like zinc finger protein (ZFP) or TAL effector (TALE), with a protein having a cleaving activity, like nuclease domain of Fok1 restriction enzyme. Each ZFP recognizes

nucleotide triplets, and a combination of 3-6 zinc fingers provides target specific binding in 9-18 base pairs (Bitinaite et al., 1998; Urnov et al., 2005). TALEs, derived from the plant pathogenic bacteria such as *Xanthomonas spp.*, contain repeat variable di-residues (RVDs) which determine the recognition of a single base pair (Boch et al., 2009; Moscou and Bogdanove, 2009). Target specific TALEs are made by a combination of 15-20 modules with containing engineered RVDs. Both ZFNs and TALENs identify the target DNA by protein-DNA interactions, so ZFP and TALE must be changed each time to match with their target sequence. For DNA cleavage by nuclease domain of Fok1 in ZFNs and TALENs, a pair of ZFNs or TALENs are required since Fok1 domain functions as a dimer (Bitinaite et al., 1998).

Unlike ZFNs and TALENs, CRISPR nucleases have different composition and mechanism. CRISPR system as a genome engineering tool is developed based on CRISPR immune system, the adaptive immune system in eubacteria and archaea (Garneau et al., 2010; Marraffini and Sontheimer, 2010). CRISPR/CRISPR-associated (Cas) immune system is divided into three stages :1) Acquisition of small foreign DNA (protospacers) from invaders in to CRISPR array; 2) Expression of *Cas* genes and transcribing trans-activating crRNA (tracrRNA) and pre-CRISPR RNA (pre-crRNA) which becomes crRNA after processing; 3) Interference of reinvading DNA, by target recognition and degradation of CRISPR components (Marraffini, 2015). CRISPR systems are classified into class 1 and class 2 (Makarova et al., 2017a, b), depending on whether it has single effector protein or multi-complexes. Class 2 system utilizes a

processed crRNA and single effector protein such as Cas9 from *Streptococcus pyogenes*, the most efficient and widely used for genome engineering in various organisms (Kim and Kim, 2014; Shmakov et al., 2017).

To cut double-stranded DNA, it requires crRNA, tracrRNA, and Cas9 protein. The crRNA is made from DNA obtained from a spacer acquisition and binds to the target DNA. The tracrRNA is complementary to some sequence of crRNA and serves to make Cas9 cleave DNA. The Cas9 protein generates DSB when the protospacer adjacent motif (PAM) is present after the crRNA binding region. For more convenient use, crRNA and tracrRNA are fused into one strand, which is called single-guide RNA (sgRNA), through the identification of essential sequence in working. By changing the sequence of crRNA or sgRNA, 20 nucleotides (nt) target sequence that hybridizes to the complementary sequence on the genome, it can easily change the cutting sites, and it has been confirmed *in vitro* and *in vivo*. Now, it is used as an easy-to-use technique, not only for knock-out research to eliminate the function of genes, but also the production of the disease model, correction of genetic disease, and improvement of animals and plants.

Modification of Cas9 protein or sgRNA, two major components of the CRISPR-Cas9 nucleases, can be used for further applications (Adli, 2018). Cas9 has two different cleavage domains, RuvC and HNH, which cleave non-target strand and target strand of DNA, respectively. Mutations in the catalytic residues of Cas9 nuclease (D10A in RuvC, H840A in HNH, or both) convert into catalytically-impaired Cas9 nickase (nCas9) or catalytically-deficient Cas9 (dCas9). nCas9 and

dCas9 still retain the property of specifically binding to the target DNA (Jinek et al., 2012), so it can be fused with various functional protein domains to induce new functions at specific genome locus. For example, CRISPR activation (CRISPRa) and CRISPR interference (CRISPRi) systems regulate the gene expression by fused domain of transcription activator, four herpes simplex virus VP16 molecules (called VP64), or transcription repressor, the Krüppel-associated box (KRAB) with dCas9 (Gilbert et al., 2014; Konermann et al., 2015; Maeder et al., 2013; Perez-Pinera et al., 2013). Also, additional RNA motif in sgRNA, linked at positions that do not affect the formation of the complex with Cas9, recruits other functional domains (Ma et al., 2018; Zalatan et al., 2015).

For therapeutic use, it is crucial to target the point mutations, a major source of human genetic disorders (Rees and Liu, 2018) and diversity. Introducing substitutions in the target locus is necessary for correction of disease or studying the function of specific nucleotides. Using CRISPR nucleases and a donor DNA template with point mutations, it is now possible to induce point mutations at the desired site. However, the efficiency of HR-induced point mutations is much lower than competitive NHEJ pathway, so there have been attempts to increase efficiency. Through NHEJ inhibitor or cell cycle regulation, HR efficiency is slightly increased, but it is not sufficient compared to the frequency of indels with NHEJ (Chu et al., 2015; Maruyama et al., 2015).

CRISPR RNA-guided programmable deaminases, composed of a nCas9 or dCas9 from *S. pyogenes* and a deaminase protein from various sources, enable

targeted nucleotide substitutions or base editing in a genome without producing DSBs. These base editing systems can be divided into two categories: cytosine base editors (CBEs) that convert a C:G base pair to a T:A pair (Hess et al., 2016; Komor et al., 2016; Ma et al., 2016; Nishida et al., 2016; Yang et al., 2016) and adenosine base editors (ABEs) that convert a A:T pair to a G:C pair (Gaudelli et al., 2017). Unlike programmable nucleases which induce small indels at a target site, programmable deaminases convert bases within a window of several nucleotides at a target site without relying on endogenous DSB repair pathways. nCas9 or dCas9 binds to target DNA and generates single-stranded DNA bubble by forming R-loop complex (Jiang et al., 2016). Then, linked single-strand-specific deaminase convert the exposed cytosine or adenine to uracil (U) or inosine (I) on the non-target strand. By DNA repair systems and replication process, U:G or I:T mismatches are resolved to produce U:A or I:C and lastly, T:A or G:C base pairs. Through these mechanisms, programmable deaminases can correct point mutations causing genetic diseases or create single-nucleotide polymorphisms (SNPs) of interest in human cells (Gaudelli et al., 2017; Kim et al., 2017a; Komor et al., 2016), animals (Kim et al., 2017b; Ryu et al., 2018), and plants (Kang et al., 2018; Zong et al., 2017).

Target specificity of CRISPR nucleases is determined by the sequence of sgRNA. The sgRNA binds to its complementary target DNA and specifies the position to cut. But there are hundreds of homologous sites that differ from the on-target sequence by up to 5-nt mismatches in the human genome; indeed, unwanted mutations caused by off-target effects of CRISPR nucleases have been reported

(Cho et al., 2014; Fu et al., 2013; Hsu et al., 2013; Kosicki et al., 2018). It is important to develop a method to identify the off-targets across the genome and to reduce them. Several methods for off-target detection are established using computational prediction tool (Bae et al., 2014; Heigwer et al., 2014; Ran et al., 2013b), capturing bound target with Cas9 (Duan et al., 2014; Kuscu et al., 2014; Singh et al., 2015; Wu et al., 2014), or capturing the consequences of DSBs (Frock et al., 2015; Kim et al., 2015; Tsai et al., 2017; Tsai et al., 2015; Wang et al., 2015) *in vitro* and *in vivo*. Also, various strategies have been reported to reduce off-target effects of Cas9 nucleases: alteration the length of sgRNA including shortening (Fu et al., 2014) or adding guanine at 5' end (Cho et al., 2014), weakening the non-specific interaction between Cas9 protein and non-target (Slaymaker et al., 2016) or target strand (Kleinstiver et al., 2016) of DNA, using as paired Cas9 nickases (Cho et al., 2014; Mali et al., 2013a; Ran et al., 2013a) or dCas9-FokI (Guilinger et al., 2014; Tsai et al., 2014), and delivery as ribonucleoprotein (Kim et al., 2014; Ramakrishna et al., 2014; Zuris et al., 2015). However, genome-wide target specificities of programmable deaminases remain unknown, owing to a lack of appropriate methods. Existing methods for identifying genome-wide off-targets are based on capturing of DSBs introduced by CRISPR nucleases, and it is not suitable for programmable deaminases.

Here, I modify Digenome-seq to assess the genome-wide specificities of CRISPR RNA-guided programmable deaminases. I identify off-targets of CBEs and confirm that CBEs are highly specific compared to CRISPR nucleases. By using modified sgRNAs, I find that both extended and truncated sgRNAs reduce

the off-target effects at identified off-target sites (Kim et al., 2017a). Furthermore, extended sgRNAs show base editing in additional nucleotides at the 5' end, demonstrating that the targeting-scope of CRISPR RNA-guided deaminases have been increased (Ryu et al., 2018).

Ⅱ. Materials and Methods

1. Cell culture and transfection conditions

HEK293T cells (ATCC CRL-11268) and NIH3T3 cells (ATCC CRL-1658) were maintained in DMEM medium supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin (Welgene) and verified using STR profile. Cells were not tested for mycoplasma contamination.

For CBE experiments, HEK293T cells (~ 1.5×10^5) were seeded on 24-well plates and transfected at ~80% confluency with Base Editor-expression plasmid (Addgene plasmid #73019, #73020, #73021) (1.5 μg) or Cas9-expression plasmid (Addgene plasmid #43945) and sgRNA-expressing plasmid (500 ng) using Lipofectamine 2000 (Thermo Fisher Scientific). NIH3T3 cells (1×10^5) were electroporated with Base Editor-expression plasmid (1.3 μg) and sgRNA plasmid (1.3 μg) via Neon Transfection System. Genomic DNA was isolated using a DNeasy Blood & Tissue Kit (Qiagen) at 72 h after transfection.

For ABE experiments, HEK293T cells (~1×10^5) were seeded on 48-well plates (Corning) and transfected at ~70% confluency with ABE7.10-expressing plasmid (750 ng) and sgRNA-expressing plasmid (250 ng) using Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's protocol. Genomic DNA was isolated using a DNeasy Blood & Tissue Kit (Qiagen) at 96 h after transfection.

2. Purification of base editor protein

The plasmid encoding the His6-rAPOBEC1-XTEN-dCas9 protein (pET28b-BE1) was a gift from David Liu (Addgene plasmid #73018). The plasmid encoding the His6-rAPOBEC1-XTEN-nCas9 protein (BE3ΔUGI) was generated by site-directed mutagenesis. Rosetta expression cells (EMD Millipore) were transformed with His6-rAPOBEC1-XTEN-nCas9 and cultured overnight in Luria-Bertani (LB) broth containing 100 μg/ml kanamycin and 34 μg/ml chloramphenicol at 37 °C. 10 ml overnight cultures of Rosetta cells containing His6-rAPOBEC1-XTEN-nCas9 were inoculated into 400 ml LB broth containing 100 μg/ml kanamycin and 34 μg/ml chloramphenicol, and cultured at 30 °C until the OD600 reached 0.5–0.6. Cells were cooled to 16 °C for 1 h, supplemented with 0.5 mM IPTG, and cultured for 14-18 h.

For protein purification, cells were harvested by centrifugation at 5,000g for 10 min at 4 °C and lysed by sonication in 5 ml lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 1 mM DTT, and 10 mM imidazole, pH 8.0) supplemented with lysozyme (Sigma) and protease inhibitor (Roche complete, EDTA-free). The soluble lysate obtained after centrifugation at 13,000 r.p.m. for 30 min at 4 °C was incubated with Ni-NTA agarose resin (Qiagen) for 1 h at 4 °C. The lysate/Ni-NTA mixture was applied to a column and washed with a buffer (50 mM NaH_2PO_4, 300 mM NaCl, and 20 mM imidazole, pH 8.0). The BE3 protein was eluted with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, and 250 mM imidazole, pH 8.0). The eluted protein was buffer exchanged with storage buffer (20 mM HEPES-KOH

(pH 7.5), 150 mM KCl, 1 mM DTT, and 20% glycerol) and concentrated with centrifugal filter units (Millipore).

3. *In vitro* transcription of sgRNAs

sgRNA was *in vitro* transcribed through run-off transcription reactions by T7 RNA polymerase (New England BioLabs). Template DNA for sgRNA contain a T7 RNA promoter sequence followed by protospacer and sgRNA scaffold sequence. sgRNA transcription templates (200 mM) were mixed with NTP (4 mM), $MgCl_2$ (14 mM), T7 RNA polymerase (500 units), RNase inhibitor (10 units) (New England BioLabs) in a reaction volume of 100 μl for overnight at 37 °C. To remove the template DNA after RNA synthesis, transcribed RNA were pre-incubated with DNase I (4 units) (New England BioLabs) and purified using RNeasy Mini Kit (Qiagen) according to the manufacturer's manual.

4. Base editor and USER treatment of a PCR amplicon

The PCR amplicon (10 μg) containing the *EMX1* site was first incubated with the purified rAPOBEC1-nCas9 protein (4 μg) and the *EMX1*-specific sgRNA (3 μg) in a reaction volume of 100 μl for 1 h at 37 °C. It was next incubated with USER (6 units) (New England BioLabs) for 30 min at 37 °C and then subjected to agarose gel electrophoresis.

5. Base editor and USER treatment of genomic DNA

Genomic DNA was purified from HEK293T cells or mouse tissue with a DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions. Genomic DNA (10 μg) was incubated with the purified rAPOBEC1-nCas9 protein (300 nM) and an sgRNA (900 nM) in a reaction volume of 500 μL for 8 h at 37 °C in a buffer (100 mM NaCl, 40 mM Tris-HCl, 10 mM $MgCl_2$, and 100 μg/ml BSA, pH 7.9). After removal of sgRNA using RNase A (50 μg/mL), uracil-containing genomic DNA was purified with a DNeasy Blood & Tissue Kit (Qiagen). Purified genomic DNA (2 μg) was incubated with USER (6 units) in a reaction volume of 100 μL at 37 °C for 3 h and, then, purified again with a DNeasy Blood & Tissue Kit (Qiagen). The target site was PCR-amplified using SUN-PCR blend and subjected to Sanger sequencing to check BE3-mediated cytosine deamination and USER-mediated DNA cleavage.

6. Whole genome sequencing and digenome sequencing

Genomic DNA (1 μg) was fragmented to the 400- to 500-bp range using the Covaris system (Life Technologies) and blunt-ended using End Repair Mix (Thermo Fischer). Fragmented DNA was ligated with adapters to produce libraries, which were then subjected to WGS using HiSeq X Ten Sequencer (Illumina) at Macrogen. WGS was performed at a sequencing depth of 30–40×. A DNA cleavage score was assigned to each nt position across the genome according to the equation.

DNA cleavage sites were identified using Digenome 1.0 and Digenome 2.0 programs. The source code of the version of Digenome 2.0 used in the manuscript is available as up-to-date versions of the program at https://github.com/chizksh/digenome-toolkit2.

For analysis of mouse WGS data, SNVs in *Tyr* #4 were trimmed out, compared with those in the WT control using the program 'Strelka' with the default 'eland' option. The resulting putative off-target sites were compared with homologous sites, identified using Cas-OFFinder, with up to seven mismatches or with up to five mismatches and a DNA or RNA bulge.

7. Targeted deep sequencing

On-target and potential off-target sites were amplified with a KAPA HiFi HotStart PCR kit (KAPA Biosystems #KK2501) for deep sequencing library generation. For the paired-end sequencing (Figure 1) (Park et al., 2017), DNA library must have two adaptors including i5 & i7 index. Target DNA was amplified using PCR polymerase with pre-index tailed primers. After pre-indexed amplification, PCR products were amplified with universal index primers (Table 1). Amplicons were purified using column-based purification kit (MGmed). Pooled PCR amplicons were sequenced using MiniSeq or Miseq with TruSeq HT Dual Index system (Illumina).

8. ABE mRNA preparation

The pET_ABE7.10_nCas9 plasmid was generated based on ABE amino acid sequences from David Liu's study (Gaudelli et al., 2017). The codons were optimized for expression in human cells. PCR using Phusion High-Fidelity DNA Polymerase (Thermo Scientific) and primers (F: 5′-GGT GAT GTC GGC GAT ATA GG-3′, R: 5′-CCC CAA GGG GTT ATG CTA GT-3′) that were described in previous paper (Kim et al., 2017b) were used to prepare the ABE7.10 mRNA template. An *in vitro* RNA transcription kit (mMESSAGE mMACHINE T7 Ultra kit, Ambion) and a purification kit, MEGAclear kit (Ambion), were used to synthesize and purify the ABE7.10 mRNA.

9. Animals

Protocols involving mice used in experiments were approved by the Institutional Animal Care and Use Committee (IACUC) of Seoul National University. C57BL/6J and ICR mouse strains were used as embryo donors and foster mothers, respectively, and C57BL/6J, ICR, and *Dmd* knockout mice were maintained in an SPF (specific pathogen-free) facility under a 12-h dark-light cycle.

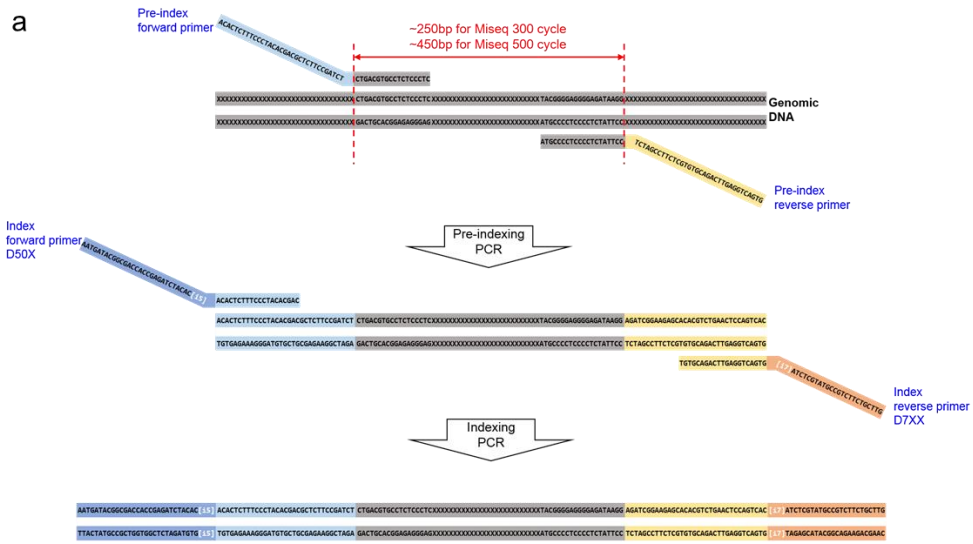
10. Microinjection of mouse zygotes

Methods for steps to prepare for microinjection, including superovulation

and embryo collection, were described in previous manuscript (Hur et al., 2016). A solution containing ABE7.10 mRNA (10 ng/μl) and sgRNA (100 ng/μl) was diluted with diethyl pyrocarbonate (DEPC)-treated injection buffer (0.25 mM EDTA, 10 mM Tris, pH 7.4) (Kim et al., 2017b; Sung et al., 2014) and injected into pronuclei using a Nikon ECLIPSE Ti micromanipulator and a FemtoJet 4i microinjector (Eppendorf). After microinjection, embryos were cultured in microdrops of KSOM+AA medium (Millipore) at 37 ℃ in a humidified atmosphere consisting of 5% CO2 in air. Two-cell-stage embryos were then transplanted into the oviducts of 0.5-dpc (days post coitus) pseudo-pregnant foster mothers to obtain offspring. For *in vitro* analysis, we cultured microinjected zygotes for 4 days to obtain blastocysts.

11. Statistical analysis

All group results are expressed as mean ± s.e.m. unless stated otherwise. Comparisons between groups were made using the two-tailed Student's t-test. Statistical significance as compared with controls was denoted with *P < 0.05, **P < 0.01, ***P < 0.001 in the figures and figure legends. Statistical analysis was performed in Graph Pad PRISM 7.



b Pre-index forward primer (5' to 3')

ACACTCTTTCCCTACACGACGCTCTTCCGATCT[gDNA target]

Pre-index reverse primer (5' to 3')

GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT[gDNA target]

c Index primers sequence

	index sequence [i5]	Index forward primer sequence
D501	tatagcct	AATGATACGGCGACCACCGAGATCTACACtatagcctACACTCTTTCCCTACACGAC
D502	atagaggc	AATGATACGGCGACCACCGAGATCTACACatagaggcACACTCTTTCCCTACACGAC
D503	cctatcct	AATGATACGGCGACCACCGAGATCTACACcctatcctACACTCTTTCCCTACACGAC
D504	ggctctga	AATGATACGGCGACCACCGAGATCTACACggctctgaACACTCTTTCCCTACACGAC
D505	aggcgaag	AATGATACGGCGACCACCGAGATCTACACaggcgaagACACTCTTTCCCTACACGAC
D506	taatctta	AATGATACGGCGACCACCGAGATCTACACtaatcttaACACTCTTTCCCTACACGAC
D507	caggacgt	AATGATACGGCGACCACCGAGATCTACACcaggacgtACACTCTTTCCCTACACGAC
D508	gtactgac	AATGATACGGCGACCACCGAGATCTACACgtactgacACACTCTTTCCCTACACGAC
	index sequence [i7]	Index reverse primer sequence
D701	cgagtaat	CAAGCAGAAGACGGCATACGAGATcgagtaatGTGACTGGAGTTCAGACGTGT
D702	tctccgga	CAAGCAGAAGACGGCATACGAGATtctccggaGTGACTGGAGTTCAGACGTGT
D703	aatgagcg	CAAGCAGAAGACGGCATACGAGATaatgagcgGTGACTGGAGTTCAGACGTGT
D704	ggaatctc	CAAGCAGAAGACGGCATACGAGATggaatctcGTGACTGGAGTTCAGACGTGT
D705	ttctgaat	CAAGCAGAAGACGGCATACGAGATttctgaatGTGACTGGAGTTCAGACGTGT
D706	acgaattc	CAAGCAGAAGACGGCATACGAGATacgaattcGTGACTGGAGTTCAGACGTGT
D707	agcttcag	CAAGCAGAAGACGGCATACGAGATagcttcagGTGACTGGAGTTCAGACGTGT
D708	gcgcatta	CAAGCAGAAGACGGCATACGAGATgcgcattaGTGACTGGAGTTCAGACGTGT
D709	catagccg	CAAGCAGAAGACGGCATACGAGATcatagccgGTGACTGGAGTTCAGACGTGT
D710	ttcgcgga	CAAGCAGAAGACGGCATACGAGATttcgcggaGTGACTGGAGTTCAGACGTGT
D711	gcgcgaga	CAAGCAGAAGACGGCATACGAGATgcgcgagaGTGACTGGAGTTCAGACGTGT
D712	ctatcgct	CAAGCAGAAGACGGCATACGAGATctatcgctGTGACTGGAGTTCAGACGTGT

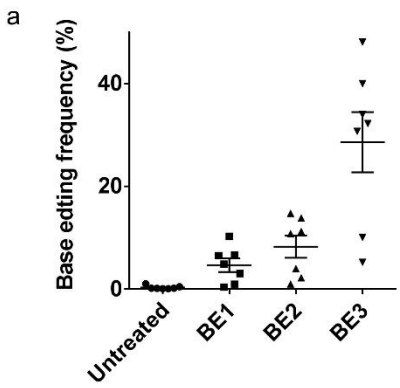
Figure 1. Illumina sequencing library preparation protocol using PCR. (a) Experimental scheme for deep sequencing library preparation. The length of target DNA ~250bp for Miseq 300 cycle or ~450bp for Miseq 500 cycle to merge the paired-end reads optimally. (b, c) List of primers used for nested indexing PCR.

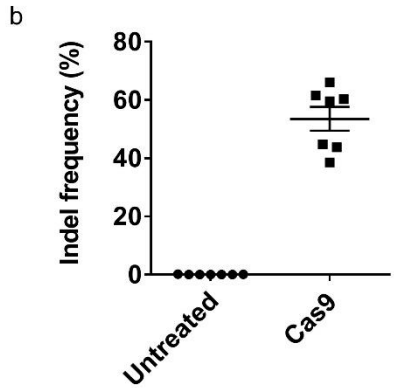
Ⅲ. Results

1. Editing efficiency and mismatch tolerance of BE3 are independent of Cas9

a. Indel efficiencies of Cas9 nucleases and base-editing efficiencies of base editors

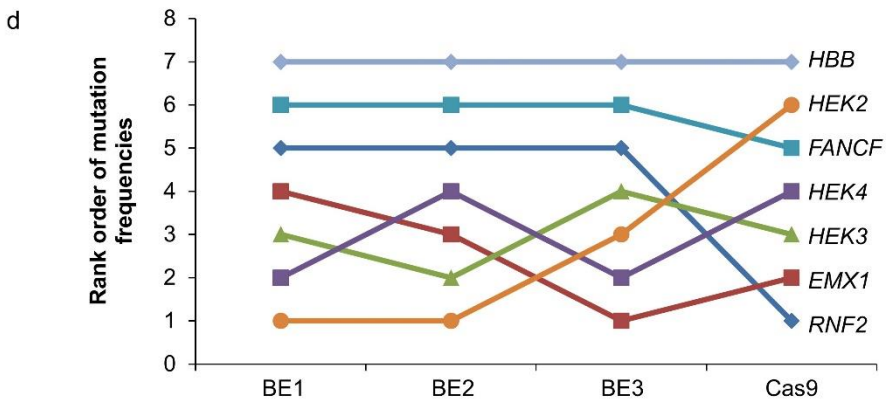
First, I compared genome editing efficiencies, defined by indel frequencies at target sites of Cas9 nucleases, with base editing efficiencies, defined by single-nucleotide substitution frequencies of three different forms of base editors (all from(Komor et al., 2016)), at seven genomic loci in HEK293T cells (Figure 2a–c). Base editors induce C-to-T (C-to-G or C-to-A, to a lesser extent) conversions at nucleotide positions 4–8 (numbered 1–20 in the 5′ to 3′ direction in protospacer sequence). Because base editor 3 (BE3: rAPOBEC1–nCas9–UGI (uracil DNA glycosylase inhibitor), 29 ± 6%) was much more efficient than BE1 (rAPOBEC1–dCas9, 5 ± 1%) and BE2 (rAPOBEC1–dCas9–UGI, 8 ± 2%), I focused on BE3 throughout this study. BE3 activities were largely independent of Cas9 nuclease activities. Thus, a rank order analysis showed that certain sgRNAs were poorly active with Cas9 but highly active with BE3, or vice versa (Figure 2d).





c

Target ID	Target site	Indel frequency (%)		Base editing frequency(%)			
		Untreated	Cas9	Untreated	BE1	BE2	BE3
EMX1	GAGTCCGAGCAGAAGAAGAA**GGG**	0.15	62	0.10	4.9	11	48
HBB	CTTGCCCCACAGGGCAGTAA**CGG**	0.069	39	0.18	0.40	1.0	5.2
RNF2	GTCATCTTAGTCATTACCTG**AGG**	0.026	66	0.17	3.0	4.0	31
FANCF	GGAATCCCTTCTGCAGCACC**TGG**	0.018	45	0.19	1.0	2.2	10
HEK2	GAACACAAAGCATAGACTGC**GGG**	0.010	44	0.087	10	15	34
HEK3	GGCCCAGACTGAGCACGTGA**TGG**	0.022	60	1.1	6.5	14	32
HEK4	GGCACTGCGGCTGGAGGTGG**GGG**	0.048	60	0.48	6.6	11	40



(With Desik Kim in Institute for Basic Science)

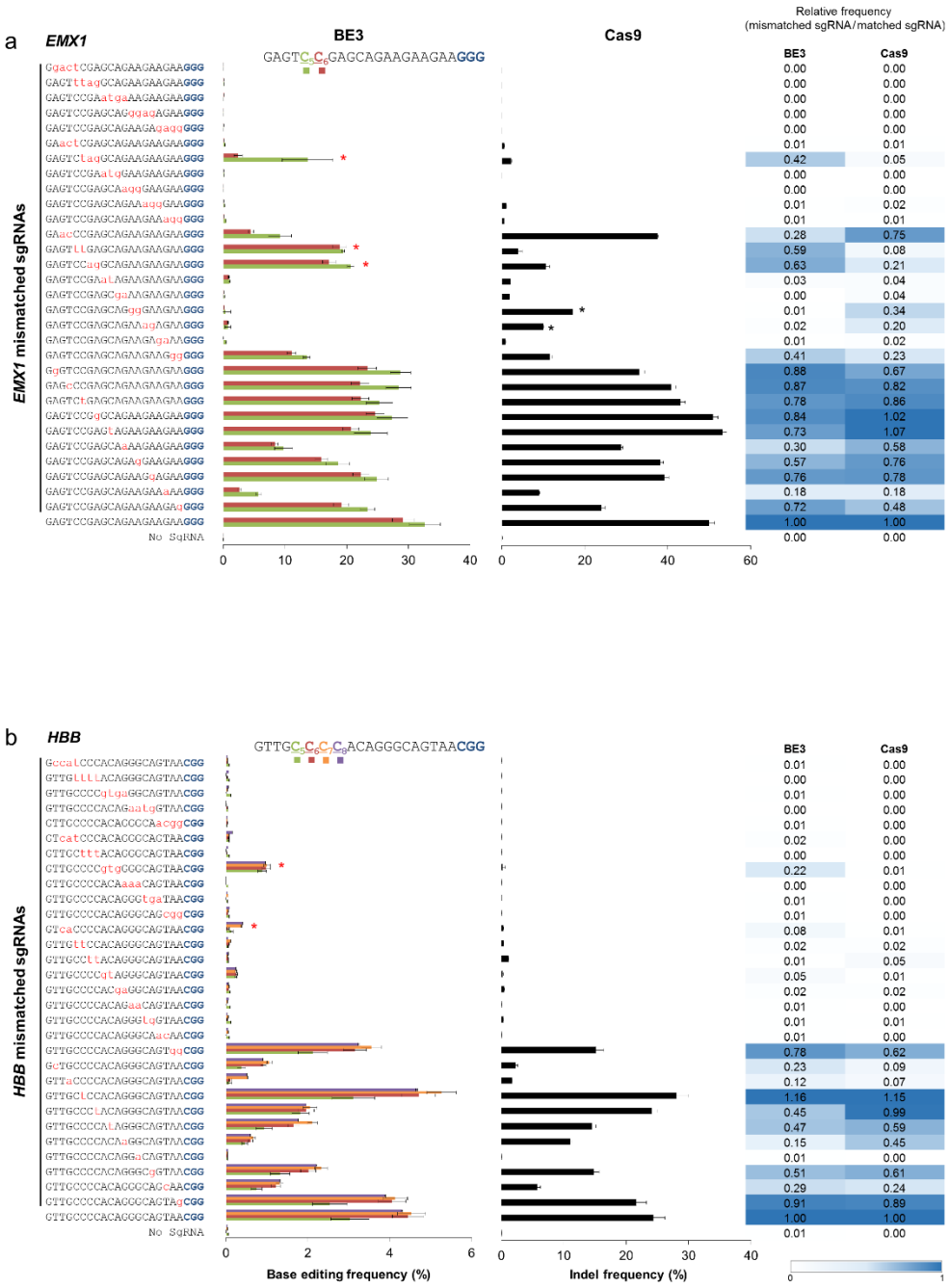
Figure 2. Comparison of BE3-associated base editing efficiencies and Cas9-associated indel frequencies in human cells. (a) Base editing efficiencies obtained with BE1 (rAPOBEC1–dCas9), BE2 (rAPOBEC1–dCas9–UGI), and BE3 (rAPOBEC1–nCas9–UGI) at seven endogenous target sites in HEK293T cells.

Base editing efficiencies were measured by targeted deep sequencing. Error bars indicate s.e.m. (b) Cas9 nuclease-driven mutation frequencies were measured by targeted deep sequencing at seven endogenous target sites in HEK293T cells. (c) A table showing target DNA sequences and mutation frequencies. The PAM is shown in blue. (d) A graph showing the rank order of indel frequencies or base editing efficiencies at seven endogenous target sites.

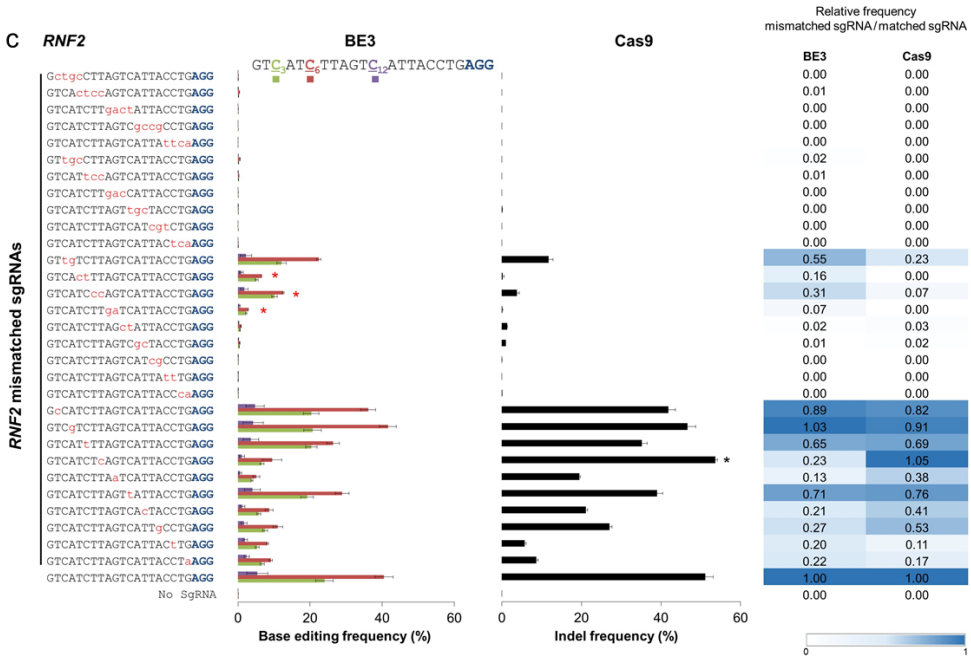
b. Tolerance of BE3 and Cas9 for mismatched sgRNAs

To assess specificities of BE3 deaminases, I examined whether BE3 can tolerate mismatches in sgRNAs. To this end, I transfected HEK293T cells with plasmids encoding BE3 or Cas9 and simultaneously with sgRNAs having one to four mismatches and measured mutation frequencies at three endogenous sites (Figure 3). Overall, there was a statistically significant correlation (R2 = 0.70, 0.83, and 0.71 at the three sites) between Cas9-induced indel frequencies and BE3-induced substitution frequencies (Figure 4). BE3 deaminases and Cas9 nucleases tolerated one-nucleotide (nt) mismatches at almost every position and 2-nt mismatches in the PAM-distal region but did not tolerate most of the 3-nt or 4-nt mismatches in either the PAM-proximal or distal regions.

Several sgRNAs (indicated by asterisks in Figure 3) with two or three mismatches were highly active with BE3 but not with Cas9 or vice versa. For example, BE3 with the fully matched sgRNA or with a 3-nt mismatched sgRNA induced substitutions at comparable frequencies (33% vs. 14%) at the *EMX1* site, whereas Cas9 with the same matched and 3-nt mismatched sgRNAs showed widely different indel frequencies (50% vs. 2%; Figure 2a). Conversely, BE3 with two 2-nt mismatched sgRNAs was poorly active (substitution frequencies <1%), whereas Cas9 with the same mismatched sgRNAs was highly active (indel frequencies >10%) (Figure 2a). These results indicate that the tolerance of Cas9 nucleases and BE3 deaminases for mismatched sgRNAs can differ and imply that BE3 and Cas9 could have separate sets of off-target sites in the genome, calling for a method to profile genome-wide specificities of RNA-programmable deaminases.

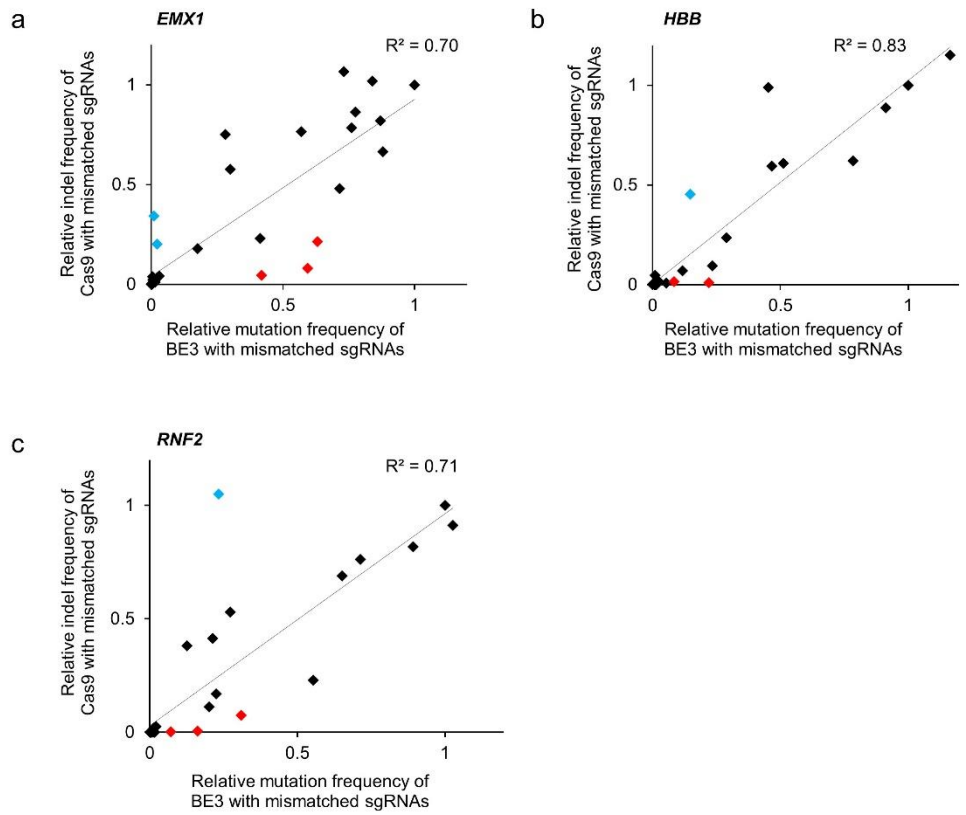

a
EMX1
BE3
Cas9
Relative frequency
(mismatched sgRNA/matched sgRNA)
EMX1 mismatched sgRNAs
No SgRNA
b
HBB
HBB mismatched sgRNAs
No SgRNA
Base editing frequency (%)
Indel frequency (%)

(*Continued*)



(By Desik Kim in Institute for Basic Science)

Figure 3. Tolerance of BE3 and Cas9 for mismatched sgRNAs. Mismatched sgRNAs that differed from the *EMX1* (a), *HBB* (b), and *RNF2* (c) target sites by 1–4 nucleotides were tested in HEK293T cells. Indel frequencies and cytosine conversion frequencies were measured using targeted deep sequencing. The PAM is shown in blue. Red or black asterisks indicate mismatched sgRNAs that were highly active with BE3 but poorly active with Cas9 or vice versa, respectively. Error bars indicate s.e.m. (n = 3).



(With Desik Kim in Institute for Basic Science)

Figure 4. Correlation between indel frequencies associated with Cas9 nucleases and base editing frequencies associated with BE3 using mismatched sgRNAs at the *EMX1* (a), *HBB* (b), and *RNF2* (c) sites. The red dots indicate mismatched sgRNAs with which the relative frequency of BE3-associated base editing was more than three times higher than the relative frequency of Cas9 nuclease-associated indels. The blue dots indicate sgRNAs with which the relative frequency of Cas9 nuclease-associated indels was more than three times higher than the relative frequency of BE3-associated base editing.

2. Digenome-seq to identify BE3 off-target sites in human the genome

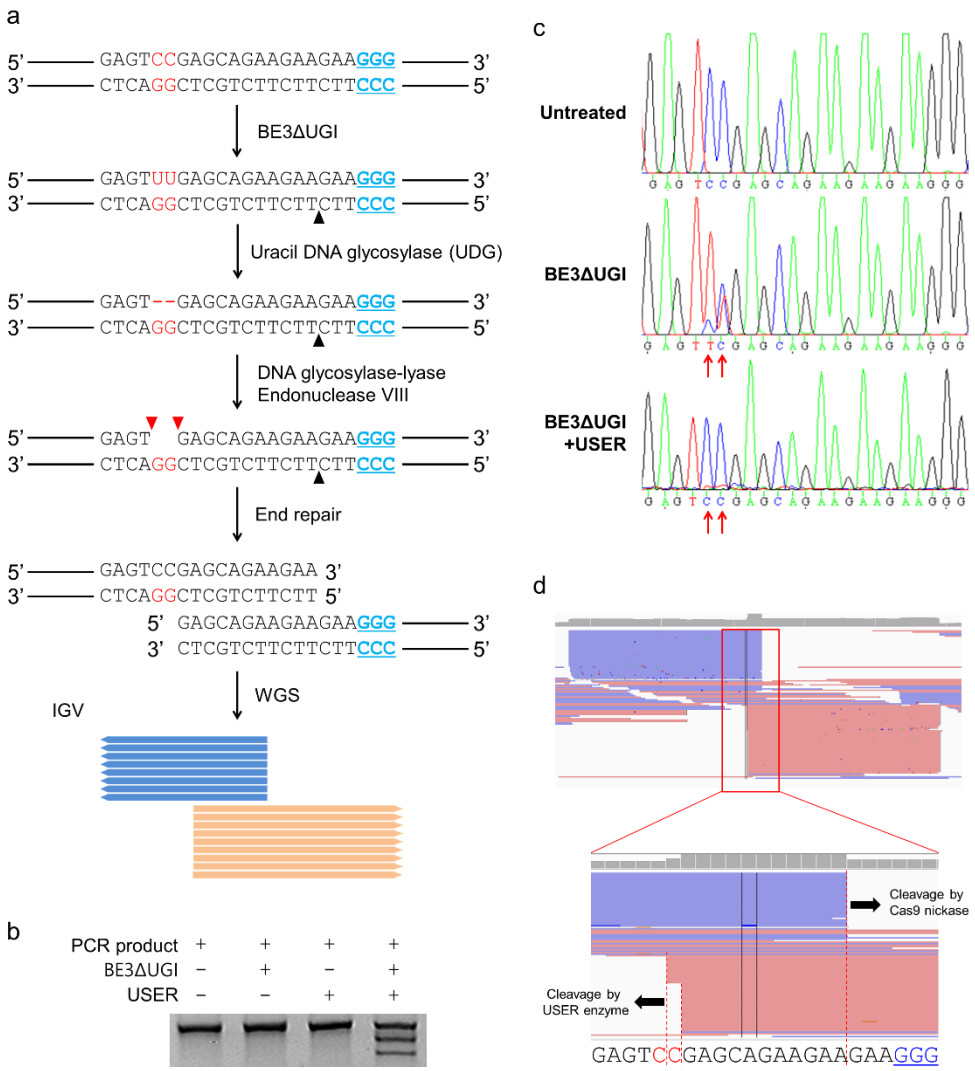
a. DSB generation of BE3-treated-DNA for Digenome-seq analysis

Several different cell-based methods, which include GUIDE-seq (Tsai et al., 2015), HTGTS (Frock et al., 2015), BLESS (Crosetto et al., 2013; Ran et al., 2015), and IDLV capture (Gabriel et al., 2011; Wang et al., 2015) have been developed and used for identifying genome-wide off-target sites at which Cas9 nucleases induce DSBs. None of these methods, at least in their present forms, are suitable for assessing the genome-wide specificities of programmable deaminases, simply because deaminases do not yield DSBs.

DSBs could be produced at deaminated, uracil-containing sites *in vitro* using appropriate enzymes and that these DNA cleavage sites could be identified via Digenome-seq, an *in vitro* method that have used for assessing genome-wide specificities of Cas9 and Cpf1 nucleases (Kim et al., 2015; Kim et al., 2016a; Kim et al., 2016b). To test this idea, first, a PCR amplicon containing a target sequence were incubated with the recombinant rAPOBEC1–nCas9 protein, a derivative of BE3 with no UGI domain (BE3ΔUGI), and its sgRNA *in vitro* to induce C-to-U conversions and a nick in the Watson and Crick strands, respectively. Then with USER (uracil-specific excision reagent), a mixture of Escherichia coli uracil DNA glycosylase (UDG) and DNA glycosylase-lyase endonuclease VIII, generate a gap at the location of the uracils, giving rise to a composite DSB (Figure 5a). Indeed,

the PCR amplicon was cleaved, when incubated with both BE3ΔUGI and USER (Figure 5b).

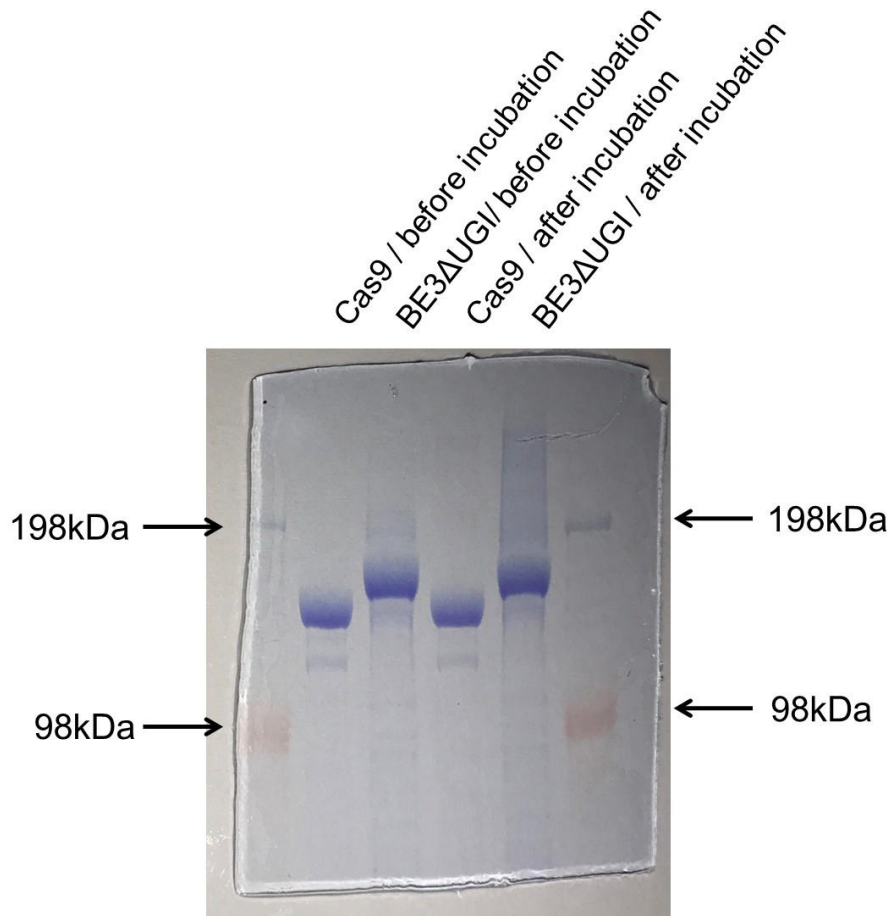
To investigate whether Digenome-seq could be used to assess genome-wide target specificities of BE3ΔUGI deaminases, human genomic DNA, purified from HEK293T cells, was incubated with each of seven BE3ΔUGI ribonucleoproteins (RNPs) (300 nM rAPOBEC1–nCas9 protein and 900 nM each of seven different sgRNAs) for 8 h (Figure 6) and then with USER for 3 h (Figure 5a). C-to-U conversions induced by BE3ΔUGI and uracil removal and DNA cleavage by USER were confirmed by quantitative real-time PCR (Figure 7) and Sanger sequencing (Figure 5c). Each genomic DNA sample was subjected to whole genome sequencing (WGS) after end repair and adaptor ligation (Figure 5a). After sequence alignment to the human reference genome (hg19), alignment patterns at each on-target site were checked using integrative genomics viewer (IGV). Uniform alignments of sequence reads, signature patterns associated with DSBs produced *in vitro*, were observed at all seven on-target sites (Figure 5d and Figure 8).



(By Desik Kim in Institute for Basic Science)

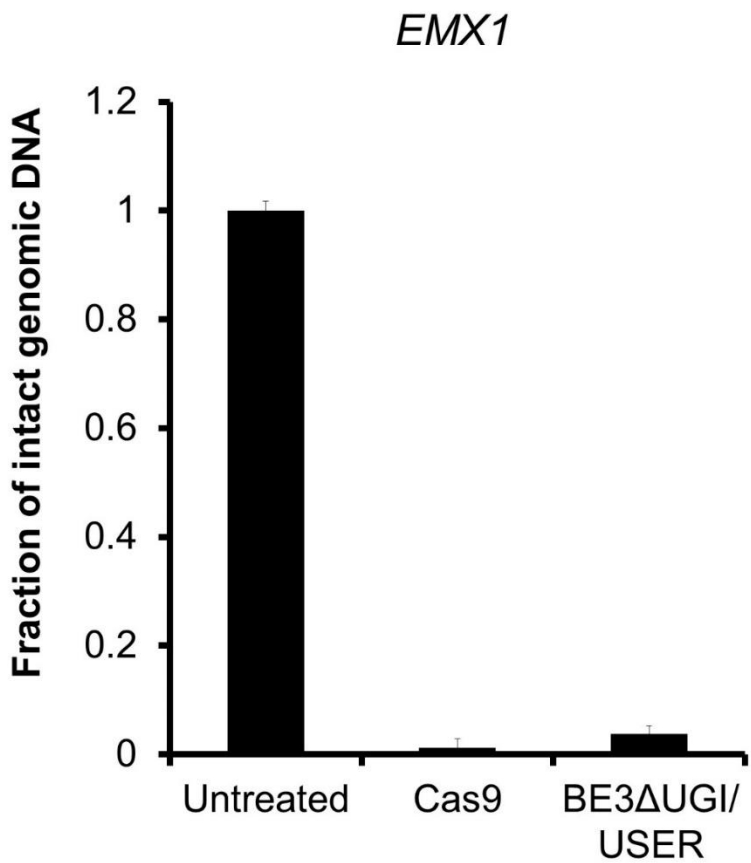
Figure 5. Digenome-seq to identify BE3 off-target sites in the human genome. (a) Overview of BE3ΔUGI Digenome-seq. BE3ΔUGI-mediated uracil-containing sites were cleaved by USER, a mixture of E. coli uracil DNA glycosylase (UDG) and DNA glycosylase-lyase endonuclease VIII. Black and red arrowheads indicate the positions of phosphodiester bonds cleaved by the Cas9 nickase and USER,

respectively. (b) The PCR product was cleaved, when treated with both BE3ΔUGI and USER. (c) Sanger sequencing results showing C-to-U conversions by BE3ΔUGI and DNA cleavage by USER. (d) IGV image showing straight alignments of sequence reads at the *EMX1* on-target site.



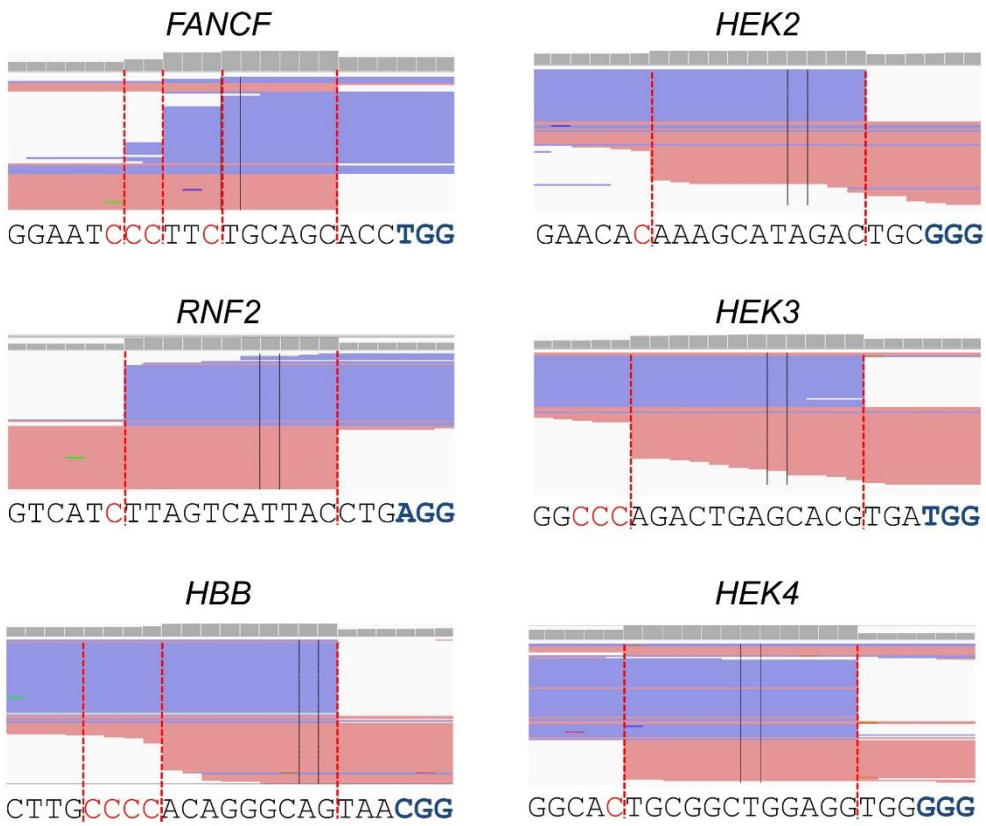
(By Desik Kim in Institute for Basic Science)

Figure 6. SDS-Polyacrylamide gel showing the integrity of Cas9 and BE3ΔUGI proteins before and after incubation with genomic DNA.



(By Desik Kim in Institute for Basic Science)

Figure 7. Quantitative real-time PCR showing almost complete cleavage of genomic DNA at the on-target site by Cas9 or BE3ΔUGI plus USER.



(By Desik Kim in Institute for Basic Science)

Figure 8. IGV images showing straight alignments of sequence reads at the 6 different on-target sites.

b. Genome-wide off-target sites of BE3 revealed by Digenome-seq

To identify BE3ΔUGI off-target sites in the human genome, DNA cleavage score (Figure 9), based on the number of sequence reads whose 5′ ends aligned at a given position, was assigned to each nt position across the genome and listed all the sites with scores >2.5, a cutoff value used for finding off-target sites of Cas9 nucleases with the same set of seven sgRNAs in previous study (Kim et al., 2016b) (Figure 10 and Table 1). Notably, only between 1 and 24 (8 ± 3) cleavage sites were observed for the seven BE3ΔUGI deaminases (plus USER) in human genomic DNA *in vitro*, far fewer than we did for Cas9 nucleases with the same set of sgRNAs (70 ± 30 sites) in a multiplex Digenome-seq analysis (Kim et al., 2016b) (Figure 11). This means that BE3ΔUGI has far fewer potential, not necessarily genuine, off-target sites than does Cas9. Sequence logos, obtained by comparing Digenome-identified sites, showed that both the PAM-distal and proximal regions contributed to the specificities of BE3ΔUGI deaminases (Figure 10c,d).

To improve the computer program (Digenome 2.0) for identifing potential off-target sites more comprehensively, the number of positions whose DNA cleavage scores over a cutoff value were counted for range from 0.0001 to 10. And the number of PAM (5′-NGN-3′ or 5′-NNG-3′)-containing sites with ten or fewer mismatches were counted, compared to the on-target site, among the positions with scores over the cutoff value (Figure 12). WGS data obtained using intact genomic DNA, which had not been treated with BE3ΔUGI and USER and thus served as a negative control, did not yield any false-positive sites with cutoff score 0.1 (Figure

12). Digenome 2.0 as cutoff score 0.1 were able to identify many additional BE3ΔUGI- and Cas9-associated DNA cleavage sites, including two sites that had been missed in previous study (Kim et al., 2016b) but had been captured by both HTGTS (Frock et al., 2015) and GUIDE-seq (Tsai et al., 2015) using the *EMX1*-specific Cas9. Using Digenome 2.0, BE3ΔUGI deaminases induced base conversions *in vitro* at 1–67 (18 ± 9) sites, whereas Cas9 nucleases cleaved genomic DNA at 30–241 (90 ± 30) sites (Figure 13).

Potential off-target sites of BE3ΔUGI- and Cas9-associated sites were carefully examined. Regardless of the number of mismatches, fewer homologous sites were identified by Digenome-seq when BE3ΔUGI was used than when Cas9 was used (Figure 14). There was a statistically significant correlation (R2 = 0.97 (score > 2.5, Digenome 1.0) or R2 = 0.86 (Digenome 2.0)) between the number of Cas9- and BE3ΔUGI-associated sites (Figure 15a,b), suggesting that sgRNAs were the primary determinants of both Cas9 and BE3ΔUGI specificities. Also, there was a strong correlation (R2 = 0.94 (Digenome 1.0) or 0.95 (Digenome 2.0)) between the number of BE3ΔUGI-associated, Digenome-captured sites and the number of homologous sites with ≤6 mismatches in the human genome (defined as "orthogonality" in ref. 7) (Figure 15c,d).

Some off-target sites were associated either with BE3ΔUGI alone or with Cas9 alone. 69% (18/26) of the sites associated with BE3ΔUGI alone had missing or extra nucleotides, compared to their respective on-target sites, producing, respectively, an RNA or DNA bulge at the DNA–gRNA interface (Lin et al., 2014) (Table 1). By contrast, these bulge-type off-target sites were rare among Cas9-

associated sites. Thus, just 4% (25/647) of sites associated with Cas9 had missing or extra nucleotides. Also, 13% (73/548) of sites associated with Cas9 alone had no cytosines at positions 13–17 counting from the PAM, the window of BE3ΔUGI-mediated deamination (Figure 16).

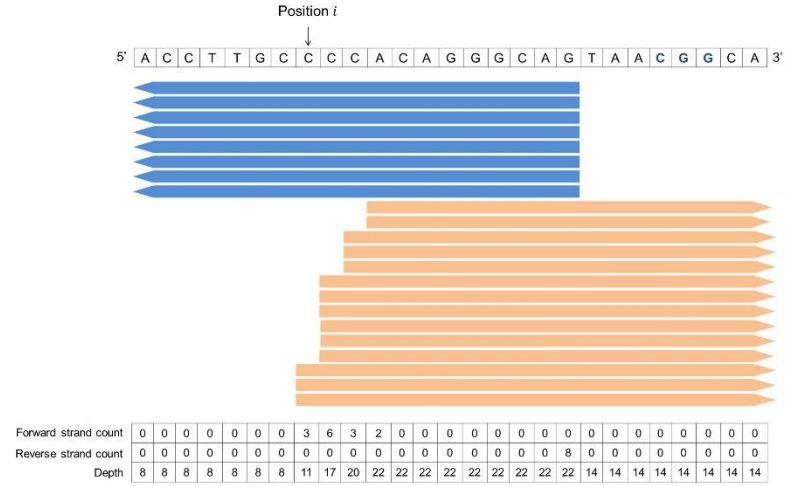
Score at position i =

$$\sum_{a=1}^{5} \frac{(F_i - 1)}{D_i} X \frac{(R_{i+8+a} - 1)}{D_{i+8+a}} \times (F_i + R_{i+8+a} - 2)$$
$$+ \sum_{a=1}^{5} \frac{(R_{i+11} - 1)}{D_{i+11}} X \frac{(F_{i-3+a} - 1)}{D_{i-3+a}} \times (R_{i+11} + F_{i-3+a} - 2)$$

F_i : Number of forward sequence reads starting at position i

R_i : Number of reverse sequence reads starting at position i

D_i : Sequencing depth at position i



Score at position i

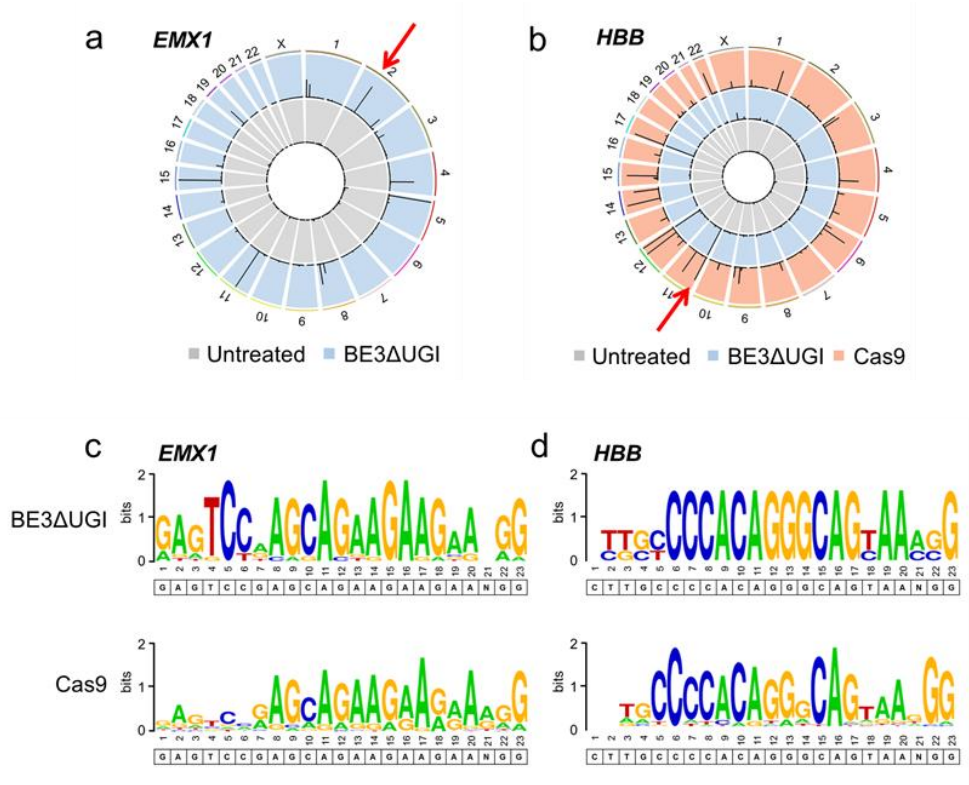
= (6-1)/17 * [(0-1)/22 * (6+0-2) + (0-1)/22 * (6+0-2) + (8-1)/22 * (6+8-2) + (0-1)/14 * (6+0-2) + (0-1)/14 * (6+0-2)]

+ (8-1)/22 * [(0-1)/8 * (8+0-2) + (3-1)/11 * (8+3-2) + (6-1)/17 * (8+6-2) + (3-1)/20 * (8+3-2) + (2-1)/22 * (8+2-2)]

= 2.66

(By Desik Kim in Institute for Basic Science)

Figure 9. *In vitro* DNA cleavage scoring system for Digenome-seq analysis of BE3ΔUGI.



(By Desik Kim in Institute for Basic Science)

Figure 10. Genome-wide BE3 off-target sites revealed by Digenome-seq. (a,b) Genome-wide circos plots representing DNA cleavage scores for *EMX1* (a) and *HBB* (b) obtained with intact genomic DNA (gray) and genomic DNA digested with BE3ΔUGI and USER (blue) or with Cas9 (red). Arrows indicate on-target sites. (c,d) Sequence logos for *EMX1* (c) and *HBB* (d) obtained via WebLogo using DNA sequences at Digenome-capture sites (DNA cleavage score > 2.5).

Table 1. Digenome-captured sites of 7 sgRNAs. PAM sequences are shown in blue. Mismatched bases are shown in red. Dashes indicate RNA bulges.

EMX1					
	Chr	Position	DNA cleavage Score	DNA seq at a cleavage sites	Bulge
EMX1_1	chr15	44109763	30.53	GAGTCtaAGCAGAAGAAGAAGAG	x
EMX1_2	chr11	62365273	26.44	GAaTCCaAGCAGAAGAAGAgAAG	x
EMX1_3	chr5	9227162	23.66	aAGTCtGAGCAcAAGAAGAATGG	x
EMX1_4	chr2	73160998	14.55	GAGTCCGAGCAGAAGAAGAAGGG	x
EMX1_5	chr4	131662222	11.14	GAaTCCaAG-AGAAGAAGAATGG	RNA bulge
EMX1_6	chr8	128801258	9.60	GAGTCCtAGCAGgAGAAGAAGAG	x
EMX1_7	chr19	24250503	8.35	GAGTCCaAGCAGtAGAgGAAGGG	x
EMX1_8	chr1	4515013	8.12	GtGTCCtAG-AGAAGAAGAAGGG	RNA bulge
EMX1_9	chr1	23720618	5.96	aAGTCCGAGgAGAgGAAGAAAGG	x
EMX1_10	chr2	219845072	5.47	GAGgCCGAGCAGAAGAAagACGG	x
EMX1_11	chr8	102244551	4.70	agtTCCaAGCAGAAGAAGcATGG	x
EMX1_12	chr3	45605387	3.11	GAGTCCacaCAGAAGAAGAAAGA	x
EMX1_13	chr16	12321159	3.01	GAGTCCaAG-AGAAGAAGtgAGG	RNA bulge
EMX1_14	chr9	111348573	1.56	GAGTCCttG-AGAAGAAGgAAGG	RNA bulge
EMX1_15	chr3	5031614	1.50	GAaTCCaAGCAGgAGAAGAAGGA	x
EMX1_16	chr14	31216733	1.34	GtacCaGAG-AGAAGAAGAgAGG	RNA bulge
EMX1_17	chr14	48932119	1.16	GAGTCCcAGCAaAAGAAGAAAAG	x
EMX1_18	chr11	107812992	1.04	aAGTCCaAGt-GAAGAAGAAAGG	RNA bulge
EMX1_19	chr12	106646090	1.03	aAGTCCatGCAGAAGAgGAAGGG	x
EMX1_20	chr2	71969823	0.80	GAGTCCtAG-AGAAGAAaAAGGG	RNA bulge
EMX1_21	chr3	145057362	0.48	GAGTCCct-CAGgAGAAGAAAGG	RNA bulge
EMX1_22	chr6	9118799	0.45	acGTCtGAGCAGAAGAAGAATGG	x
EMX1_23	chr1	59750259	0.27	GAGTtCcAGaAGAAGAAGAAGAG	x
EMX1_24	chr11	79484079	0.22	GAGTCCtAa-AGAAGAAGcAGGG	RNA bulge
EMX1_25	chr9	135663403	0.21	cAGTCCaAaCAGAAGAgGAATGG	x

HBB					
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Bulge
HBB_1	chr11	5248214	17.68	CTTGCCCCACAGGGCAGTAACGG	x
HBB_2	chr17	8370252	13.64	tTgctCCCACAGGGCAGTAAACG	x
HBB_3	chr12	124803834	10.88	gcTGCCCCACAGGGCAGcAAAGG	x
HBB_4	chrX	75006256	2.34	gTgGCCCCACAGGGCAGgAATGG	x
HBB_5	chr12	93549201	0.55	aTTGCCCCACgGGGCAGTgACGG	x
HBB_6	chr10	95791920	0.27	acTctCCCACAaGGCAGTAAGGG	x
HBB_7	chr9	104595883	0.18	tcaGCCCCACAGGGCAGTAAGGG	x

HEK2					
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Bulge
HEK2_1	chr4	90522183	18.27	GAACACAAtGCATAGAtTGCCGG	x
HEK2_2	chr5	87240613	7.54	GAACACAAAGCATAGACTGCGGG	x
HEK2_3	chr2	19844956	0.93	aActcCAAAGCATAtACTGCTGG	x

(*Continued*)

RNF2					
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Bulge
RNF2_1	chr1	185056773	27.66	GTCATCTTAGTCATTACCTGAGG	x

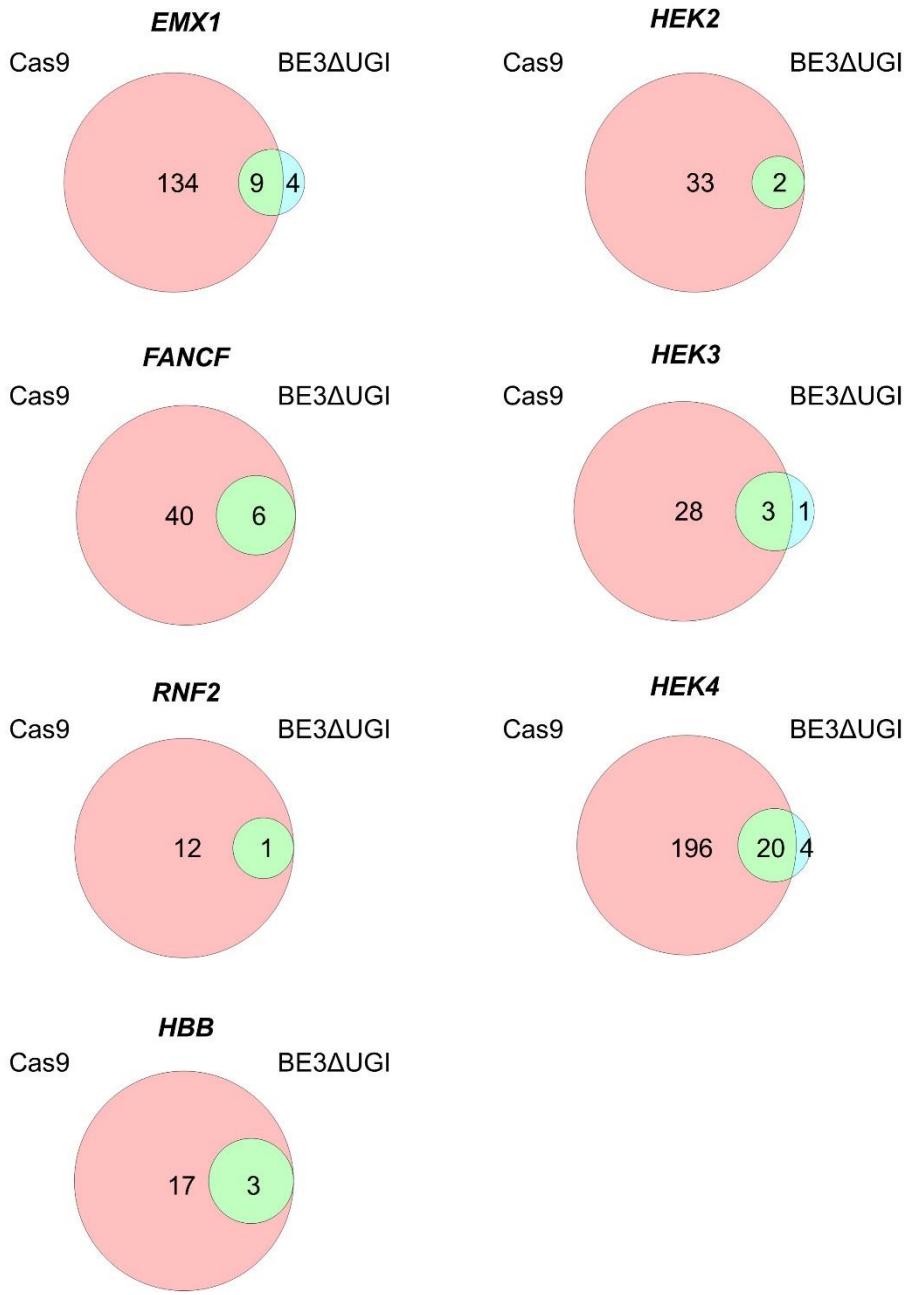
FANCF					
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Bulge
FANCF_1	chr10	73463135	13.34	tGAATCCCaTCTcCAGCACCAGG	x
FANCF_2	chr11	22647338	7.04	GGAATCCCTTCTGCAGCACCTGG	x
FANCF_3	chr10	43410030	6.53	GGAgTCCCTcCTaCAGCACCAGG	x
FANCF_4	chr10	37953199	5.67	GGAgTCCCTcCTaCAGCACCAGG	x
FANCF_5	chr11	47554037	5.13	GGAATCCCTTCTaCAGCAtCCTG	x
FANCF_6	chr16	49671025	3.00	GGAgTCCCTcCTGCAGCACCTGA	x
FANCF_7	chr18	8707528	1.26	GGAAcCCCgTCTGCAGCACCAGG	x
FANCF_8	chr7	44076496	0.95	GtctcCCCTTCTGCAGCACCAGG	x
FANCF_9	chr9	113162294	0.46	aaAATCCCTTCcGCAGCACCTAG	x
FANCF_10	chr15	49119756	0.42	tGtATttCTTCTGCctCAggCTG	x
FANCF_11	chr2	54853314	0.39	GGAATatCTTCTGCAGCcCCAGG	x
FANCF_12	chr8	21374810	0.37	GagtgCCCTgaaGCctCAgCTGG	x
FANCF_13	chrX	86355179	0.35	accATCCCTcCTGCAGCACCAGG	x
FANCF_14	chr3	35113165	0.20	tGAATCCtaaCTGCAGCACCAGG	x
FANCF_15	chr10	3151994	0.13	ctctgtCCTTCTGCAGCACCTGG	x

HEK3					
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Bulge
HEK3_1	chr1	47005705	29.27	aGCtCAGACTGAGCAaGTGAGGG	x
HEK3_2	chr9	110184636	11.38	GGCCCAGACTGAGCACGTGATGG	x
HEK3_3	chr19	882560	10.90	GGCCCAGA--GAGCACGTGtGGG	RNA bulge
HEK3_4	chr15	79749930	3.03	caCCCAGACTGAGCACGTGcTGG	x
HEK3_5	chr17	34954539	2.10	GGCCCa-ACTGAGCAaGTGATGG	RNA bulge
HEK3_6	chrX	114764149	1.66	aGaCCAGACTGAGCAaGaGAGGG	x
HEK3_7	chr6	73097166	0.15	GGCCactcaTGgcCACaTacTGG	x

HEK4					
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Bulge
HEK4_1	chr20	31349772	19.26	GGCACTGCGGCTGGAGGTGGGGG	x
HEK4_2	chr6	160517881	15.45	GGCACTGCtGCTGGgGGTGGTGG	x
HEK4_3	chr6	168787137	15.37	GGCACTGCa-CTGGAGGTtGTGG	RNA bulge
HEK4_4	chr19	33382081	13.83	GGCtCTGCGGCTGGAGGgGGTGG	x
HEK4_5	chr20	60080553	12.71	aGCACTGCaGaTGGAGGaGGCGG	x
HEK4_6	chr5	141232853	10.87	GGCACTGCGGCaGGgaGgaGGGG	x
HEK4_7	chr20	60010562	10.51	tGCACTGCGGCcGGAGGaGGTGG	x
HEK4_8	chr13	70136736	8.76	GGCACT-gGGCTGaAGGTaGAGG	RNA bulge
HEK4_9	chr20	1151854	8.41	GGCACTGtGGCTGcAGGTGGAGG	x
HEK4_10	chr15	71686928	7.70	tGCtCTGCGGCaGGAGGaGGAGG	x
HEK4_11	chr7	1397398	6.71	aGCACTGCaGCTGGgaGTGGAGG	x
HEK4_12	chr20	45343010	6.57	GGCACTGaGGgTGGAGGTGGGGG	x
HEK4_13	chr8	20854500	5.57	GGCACTGgGGCTGGAGacGGGGG	x
HEK4_14	chr7	54561437	5.40	aGgACTGCGGCTGGgGGTGGTGG	x
HEK4_15	chr15	60790561	5.29	GGCACTGCaaCTGGAaGTGaTGG	x

(*Continued*)

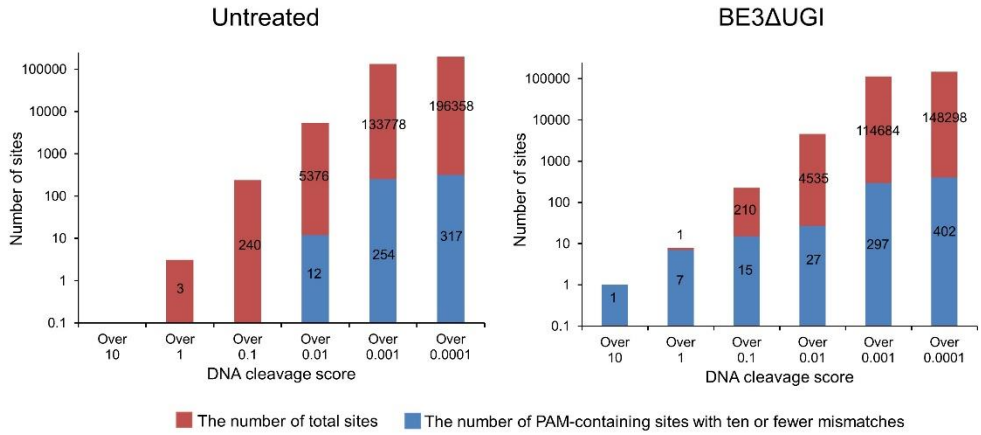
HEK4					
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Bulge
HEK4_16	chr13	27629410	4.40	GGCACTGgGGtTGGAGGTGGGGG	x
HEK4_17	chr7	110143150	3.69	GcCACTGCaGCTaGAGGTGGAGG	x
HEK4_18	chr7	139244406	3.59	GcCACTGCGaCTGGAGGaGGGGG	x
HEK4_19	chr19	2474643	3.56	GGCACTG-GGCTGGAGGcGGGGG	RNA bulge
HEK4_20	chr2	6961255	3.17	aGCtCTGCGGCaGGAGtTGGAGG	x
HEK4_21	chr17	75429280	2.90	GaCACcaCGGCTGGAGaTGGTGG	x
HEK4_22	chr7	17979717	2.66	GcactgGCaGCcGGAGGTGGTGG	DNA bulge
HEK4_23	chr9	5020590	2.64	tGCACTGCaGCTGcAGGTGGAGG	x
HEK4_24	chrX	122479548	2.52	GGCACTG-GGCTGGAGaTGGAGG	RNA bulge
HEK4_25	chr12	104739608	2.48	ccttCTGCGGCTGGAaGTGGTGG	x
HEK4_26	chr17	40693638	2.38	GcactgcaGGCaGGAGGTGaGTG	DNA bulge
HEK4_27	chr8	144781301	2.38	GaCACTGCaGCTGGAGGTGGGGT	x
HEK4_28	chr9	74103955	2.36	GGCACTGCaGCaGGgGaTGGGGG	x
HEK4_29	chr18	37194558	2.31	GGCACTGCGGgTGGAGGcGGGGG	x
HEK4_30	chr20	60895671	2.12	GGCACaGCaGCTGGAGGTGcTGG	x
HEK4_31	chr12	113935460	1.63	GGCcCTGCGGCTGGAGaTatGGG	x
HEK4_32	chrX	70597642	1.57	GaCACTGC-tCTGGAGGTGGTGG	RNA bulge
HEK4_33	chr15	41044242	1.31	GGCgCTGCGGCgGGAGGTGGAGG	x
HEK4_34	chr17	176302	1.18	tGCACTGtGGCTGGAGaTGGGGG	x
HEK4_35	chr10	77103119	1.15	GGCAtcaCGGCTGGAGGTGGAGG	x
HEK4_36	chr7	134872032	0.93	aGCACTGtGGCTGGgGGaGGCGG	x
HEK4_37	chr9	133039175	0.86	GtCACTGCaGCTGGAGGaGGGGG	x
HEK4_38	chr10	73435248	0.79	GtaACTGCGGCTGGcGGTGGTGG	x
HEK4_39	chr14	21993455	0.78	GGtACaGCGGCTGGgGGaGGCGG	x
HEK4_40	chr17	29815563	0.59	GGCgCTGCGGCcGGAGGTGGGGC	x
HEK4_41	chr16	50300346	0.56	aGCACTGtGGCTGGgGGaGGGGG	x
HEK4_42	chr11	78127584	0.53	tGCACTGCaGCTGGAGGcaaCGG	x
HEK4_43	chr19	1295086	0.52	GaCACTGaGGCaGGAGGTGGGGG	x
HEK4_44	chr2	162283033	0.51	GGCAtctgGGCTGGgGGTaGGGG	x
HEK4_45	chr20	24376056	0.47	GGCACTGaGaCcaGAGGTGGTGG	x
HEK4_46	chr16	1029977	0.42	GGCACTGCaGacGGAGGTGtGGG	x
HEK4_47	chr19	47503406	0.39	GGCACTG-GGCTGGAGGgGaGAG	RNA bulge
HEK4_48	chr2	231467380	0.39	GGCACTGCaGCTGGgGGTtGGTG	x
HEK4_49	chr10	13692636	0.38	GGCACTGgGGCTGGgGGaGGGGG	x
HEK4_50	chr1	32471659	0.34	GGCACTtCaGCTGGAGGcaGAGG	x
HEK4_51	chr17	8634933	0.33	GGCACat-GGaTGGAGGTGGAGG	RNA bulge
HEK4_52	chr6	83388605	0.30	aGCACTGtGG-TGGAGGTGGAGG	RNA bulge
HEK4_53	chr10	27700491	0.29	GGCACTG-GGtTGGgGGTGGTGG	RNA bulge
HEK4_54	chr1	143662284	0.27	GGCACat-GGCTGGgGGTGGTGG	RNA bulge
HEK4_55	chr16	49777696	0.22	tGCACTGCGaCTGGAGGgaGAGG	x
HEK4_56	chr19	38616186	0.19	GGCACTGaGaCTGGgGGTGGGGG	x
HEK4_57	chr10	126752487	0.18	GGCACTGCaGCctGgGGgtGGGG	x
HEK4_58	chr16	28266968	0.17	GGCtCTtCGGCTGGAGGTaGCGG	x
HEK4_59	chr2	149886210	0.15	GaCACTG-GGCTGGAGGTtGCGG	RNA bulge
HEK4_60	chr20	37471343	0.15	aGCACTGtGcCTGGgGGTGGGGG	x
HEK4_61	chr12	53453556	0.13	tGgACTGCGGCTGGAGagGGAGG	x
HEK4_62	chr15	30501337	0.13	GGCACTG-GGCTGGAtGTGGTGG	RNA bulge
HEK4_63	chr5	139284047	0.12	GGCACTGaGGCTGcAGGcGGCGG	x
HEK4_64	chr8	119227145	0.12	GGCACaatGGCTGGAGGTGaAGG	x
HEK4_65	chr14	95761249	0.11	GGCACTctGGCTGGAGcTGGGGG	x
HEK4_66	chr3	23651529	0.11	GGCACaGCaGgTGGAGGTGGAGG	x
HEK4_67	chr12	9287415	0.10	GGCtCTGCaGCcaGgGGTGGAGG	x



(By Desik Kim in Institute for Basic Science)

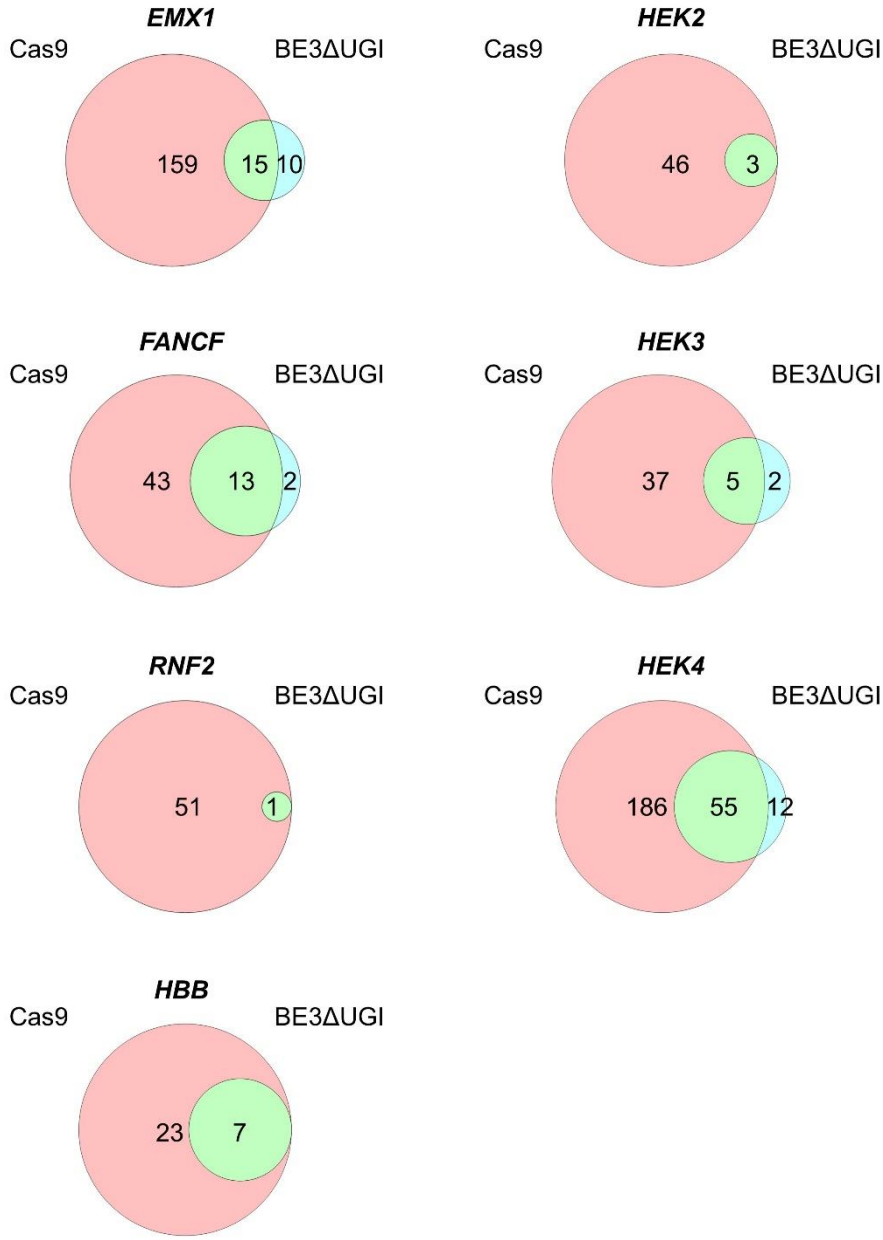
Figure 11. Venn diagrams showing the number of sites with DNA cleavage scores over 2.5 identified by Digenome-seq of Cas9 nuclease- and BE3ΔUGI-treated genomic DNA.

FANCF



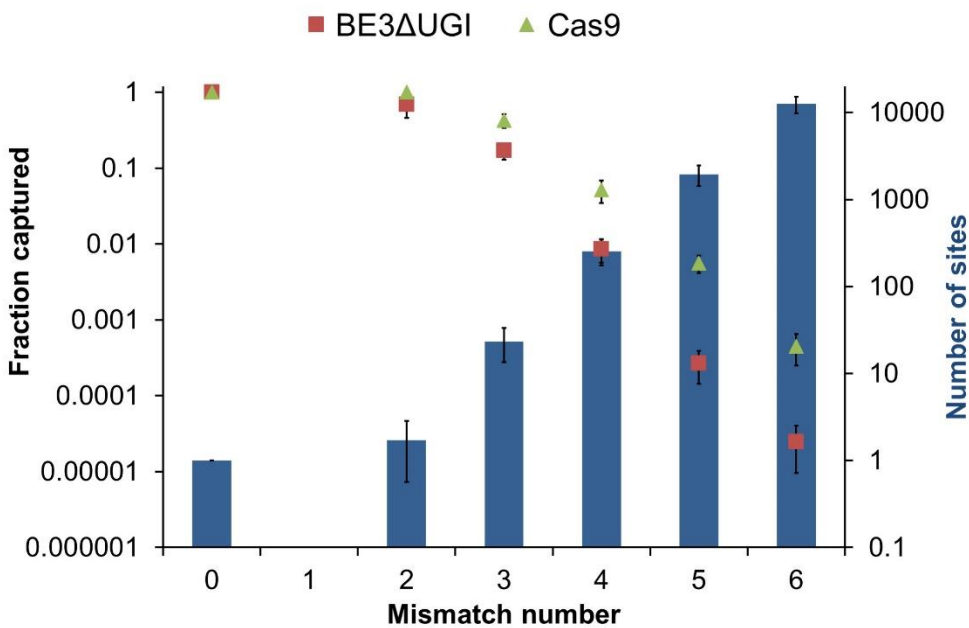
(By Desik Kim in Institute for Basic Science)

Figure 12. The number of total sites (upper) and the number of PAM-containing sites with ten or fewer mismatches (lower) for a range of DNA cleavage scores. Intact human genomic DNA (left) and genomic DNA digested by BE3ΔUGI and USER (right) were subjected to whole genome sequencing.



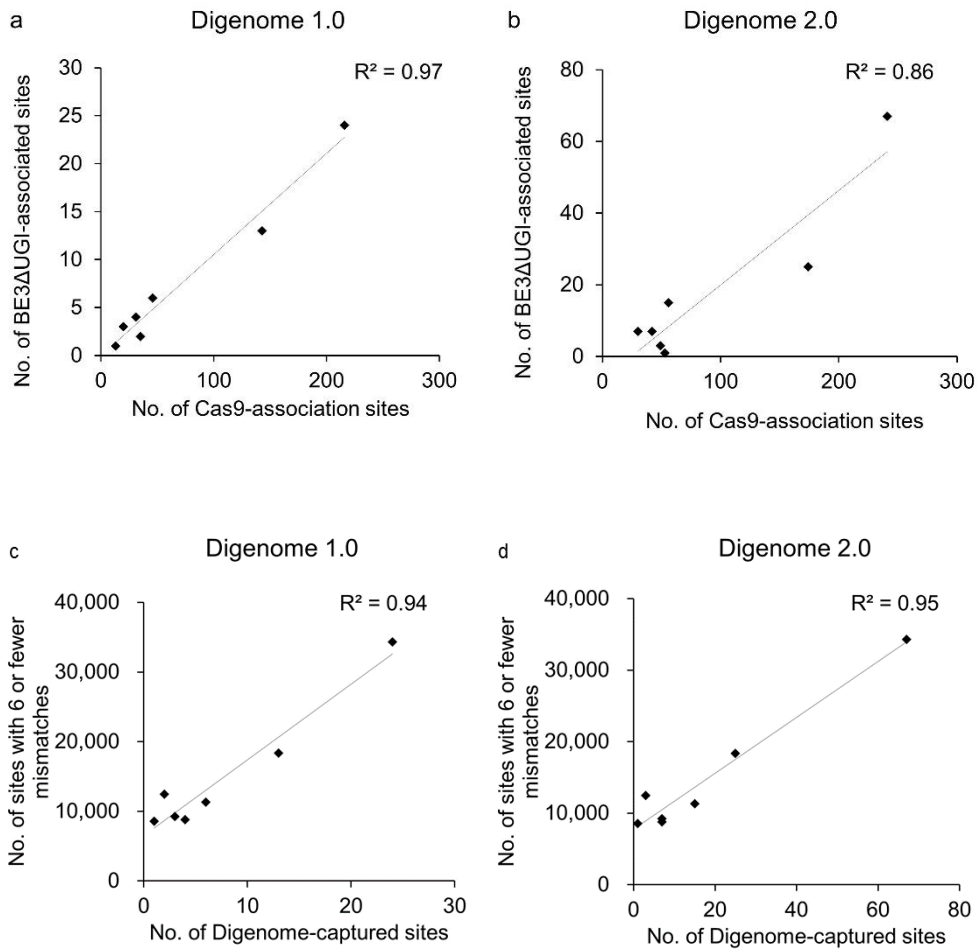
(By Desik Kim in Institute for Basic Science)

Figure 13. Venn diagrams showing the number of PAM-containing homologous sites with DNA cleavage scores over 0.1 identified by Digenome-seq of Cas9- and BE3ΔUGI-treated genomic DNA.



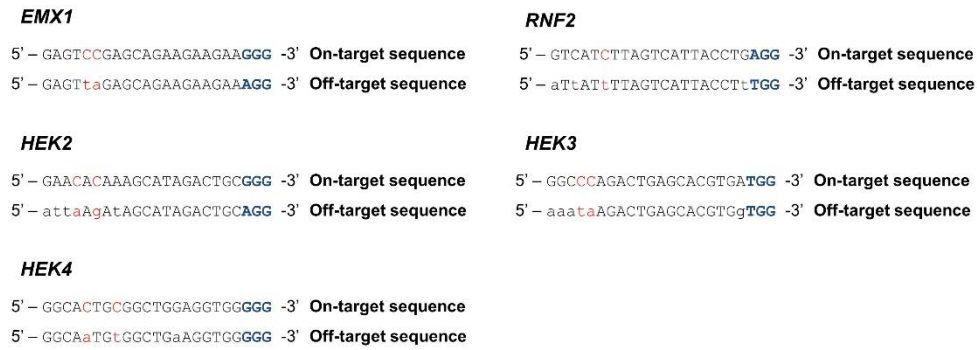
(By Desik Kim in Institute for Basic Science)

Figure 14. Fraction of homologous sites captured by Digenome-seq. Blue bars represent the number of homologous sites that differ from on-target sites by up to 6 nt. Red squares (BE3ΔUGI) and green triangles (Cas9) represent the fraction of Digenome-identified sites for a range of mismatch numbers.



(By Desik Kim in Institute for Basic Science)

Figure 15. Analysis of correlations revealed by Digenome 1.0 and Digenome 2.0. (a, b) Correlation between the number of BE3ΔUGI- and Cas9-associated sites identified by Digenome 1.0 (a) and Digenome 2.0 (b). (c, d) Correlation between the number of BE3ΔUGI-associated sites identified by Digenome 1.0 (c) or Digenome 2.0 (d) and the number of sites with 6 or fewer mismatches.



(With Desik Kim in Institute for Basic Science)

Figure 16. Examples of Digenome-captured off-target sites associated only with Cas9, which contain no cytosines at positions 13-17. The PAM is shown in blue. Mismatched bases are shown in red.

3. Validation of genome-wide BE3 off-target sites

To validate off-target effects at BE3ΔUGI-associated sites identified by Digenome-seq, I performed targeted deep sequencing and measured BE3-induced substitution frequencies and Cas9-induced indel frequencies in HEK293T cells (Figure 17 and Table 2). I analyzed a total of 75 sites identified using seven sgRNAs and observed BE3-induced point mutations at 50 sites, including all seven on-target sites, with frequencies above noise levels caused by sequencing errors (typically in the range of 0.1~1%), resulting in a validation rate of 67%. It is possible that BE3 can still induce mutagenesis at the other BE3ΔUGI-associated, Digenome-captured sites with frequencies below background noise levels. Notably, I was able to identify BE3 off-target sites at which base editing was detected with a frequency of 0.1%, demonstrating that Digenome-seq is a highly sensitive method. Cas9 nucleases detectably induced indels at 70% (44/63) of the sites associated with both Cas9 and BE3ΔUGI but failed to do so at each of the 12 sites associated with BE3ΔUGI alone (Table 2). Likewise, BE3 did not cause detectable substitutions at 24 Digenome-captured sites that were associated with Cas9 nucleases alone using three different sgRNAs (Table 3). Furthermore, six BE3s did not induce base editing at 57 Digenome-negative sites with ≤3 mismatches, identified using Cas-OFFinder (Bae et al., 2014) (Table 4).

Frequencies of BE3-induced substitutions were well-correlated with those of Cas9-mediated indels (R2 = 0.92 (*EMX1*) or 0.89 (*HBB*)) (Figure 17a,b). Nevertheless, there were many off-target sites validated in the BE3 but not by Cas9 data. 64% (7/11) of these validated, BE3-exclusive off-target sites had a missing

nucleotide, compared to their respective on-target sites. These results show that Cas9 and BE3 off-target sites largely overlapped with each other but that there were off-target sites exclusively associated with Cas9 alone or BE3 alone (Figure 17a,b and Table 2). This discrepancy can be attributed to the fact that BE3 and Cas9 catalyze different chemical reactions and induce site-specific mutations via different repair mechanisms.

To compare BE3 off-target effects and Cas9 off-target effects quantitatively, an off-target effect index (OTI) (Kim et al., 2016b) for each of the seven sgRNAs was obtained by dividing cumulative mutation frequencies at validated off-target sites by the on-target mutation frequency (Table 5). BE3 OTIs, which ranged from 0.0 to 1.7, were always smaller than Cas9 OTIs, suggesting that BE3 deaminases are more specific than Cas9 nucleases in general.

In addition to human genomic DNA, Digenome-seq using mouse genomic DNA were carried out to test whether the BE3ΔUGI deaminase targeted to the *Dmd* gene to create mutant mice with nonsense or missense mutations (Kim et al., 2017b) was highly specific in the mouse genome (Figure 18a). Two candidate sites with two mismatches and a missing nucleotide were identified in addition to the on-target site but were invalidated by targeted deep sequencing (Figure 18b,c). Thus, no point mutations were detectably induced at these two sites in a mouse cell line, resulting in an OTI of 0.0.

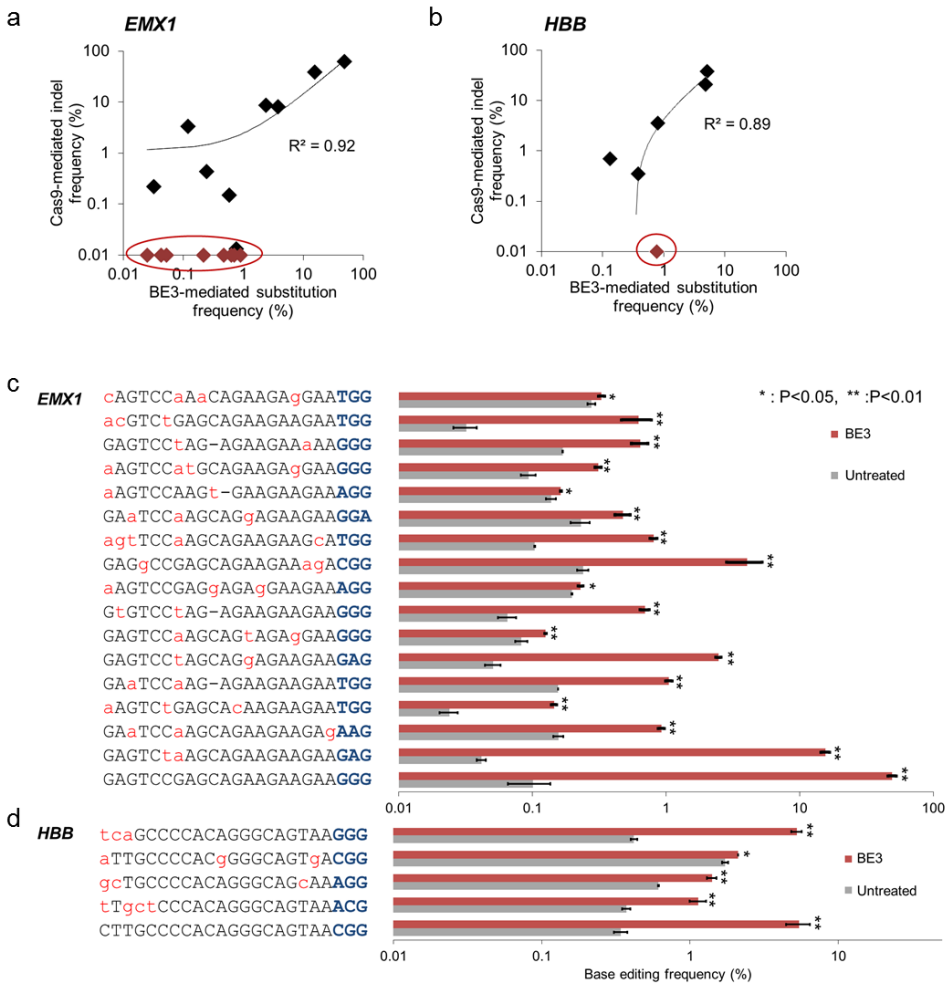


Figure 17. Validation of genome-wide BE3 off-target sites revealed by Digenome-seq. (a,b) Scatterplots for *EMX1* (a) and *HBB* (b) of BE3-mediated substitution frequencies vs. Cas9-mediated indel frequencies determined using targeted deep sequencing. Circled red dots indicate off-target sites validated by BE3 but invalidated by Cas9. (c,d) BE3 off-target sites for *EMX1* (c) and *HBB* (d) validated in HEK293T cells by targeted deep sequencing. PAM sequences are shown in blue. Mismatched bases are shown in red. Dashes indicate RNA bulges. Error bars indicate s.e.m. (n = 3).

Table 2. Mutation frequencies of Cas9 and BE3 at on-target and off-target sites captured by Digenome-seq.

		EMX1																										
		Base editing efficiency, C to other bases (%)																							Indel frequency (%)		Validation	
																									(-) RGEN	RGEN	BE3	Cas9
On-target (EMX1_4)		G	A	G	T	C	C	G	A	G	C	A	G	A	A	G	A	A	G	A	A	G	G	G	0.15	61.59	Validated	Validated
	Untreated					0.04	0.06				0.15																	
	BE1					8.49	4.72				0.08																	
	BE2					11.08	10.72				0.09																	
	BE3					49.17	45.06				0.10																	
EMX1_1		G	A	G	T	C	t	a	A	G	C	A	G	A	A	G	A	A	G	A	A	G	A	G	0.29	38.25	Validated	Validated
	Untreated					0.04					0.05																	
	BE1					3.13					0.05																	
	BE2					0.75					0.05																	
	BE3					15.57					0.07																	
EMX1_2		G	A	a	T	C	C	a	A	G	C	A	G	A	A	G	A	A	G	A	g	A	A	G	0.00	0.01	Validated	Validated
	Untreated					0.08	0.08				0.07																	
	BE1					0.65	0.31				0.06																	
	BE2					0.32	0.32				0.07																	
	BE3					0.84	0.81				0.07																	
EMX1_3		a	A	G	T	C	t	G	A	G	C	A	c	A	A	G	A	A	G	A	A	T	G	G	0.10	3.45	Validated	Validated
	Untreated					0.02					0.07		0.06															
	BE1					0.02					0.07		0.04															
	BE2					0.02					0.05		0.05															
	BE3					0.13					0.07		0.05															
EMX1_5		G	A	a	T	C	C	a	A	G	-	A	G	A	A	G	A	A	G	A	A	T	G	G	0.01	0.01	Validated	Invalidated
	Untreated					0.06	0.10																					
	BE1					0.63	0.24																					
	BE2					0.32	0.34																					
	BE3					0.96	0.96																					
EMX1_6		G	A	G	T	C	C	t	A	G	C	A	G	g	A	G	A	A	G	A	A	G	A	G	0.00	8.63	Validated	Validated
	Untreated					0.02	0.04				0.04																	
	BE1					0.06	0.07				0.07																	
	BE2					0.07	0.08				0.05																	
	BE3					2.43	2.40				0.04																	
EMX1_7		G	A	G	T	C	C	a	A	G	C	A	G	t	A	G	A	g	G	A	A	G	G	G	0.01	0.01	Validated	Invalidated
	Untreated					0.03	0.06				0.06																	
	BE1					0.07	0.10				0.07																	
	BE2					0.03	0.06				0.09																	
	BE3					0.05	0.09				0.07																	
EMX1_8		G	t	G	T	C	C	t	A	G	-	A	G	A	A	G	A	A	G	A	A	G	G	G	0.08	0.08	Validated	Invalidated
	Untreated					0.05	0.03																					
	BE1					0.64	0.57																					
	BE2					0.54	0.39																					
	BE3					0.37	0.34																					
EMX1_9		a	A	G	T	C	C	G	A	G	g	A	G	A	g	G	A	A	G	A	A	A	G	G	0.01	0.23	Validated	Validated
	Untreated					0.05	0.16																					
	BE1					0.06	0.18																					
	BE2					0.06	0.17																					
	BE3					0.09	0.25																					
EMX1_10		G	A	G	g	C	C	G	A	G	C	A	G	A	A	G	A	A	a	g	A	C	G	G	0.00	7.94	Validated	Validated
	Untreated					0.14	0.10				0.13																	
	BE1					0.44	0.24				0.16																	
	BE2					0.51	0.48				0.15																	
	BE3					3.45	3.70				0.17																	
EMX1_11		a	g	t	T	C	C	a	A	G	C	A	G	A	A	G	A	A	G	c	A	T	G	G	0.00	0.01	Validated	Invalidated
	Untreated					0.06	0.05				0.07									0.06								
	BE1					1.19	0.44				0.08									0.07								
	BE2					0.46	0.43				0.05									0.07								
	BE3					0.74	0.62				0.06									0.07								
EMX1_12		G	A	G	T	C	C	a	c	a	C	A	G	A	A	G	A	A	G	A	A	A	G	A	0.00	0.00	Invalidated	Invalidated
	Untreated					0.08	0.26		0.11		0.11																	
	BE1					0.08	0.24		0.11		0.11																	
	BE2					0.08	0.23		0.10		0.10																	
	BE3					0.17	0.33		0.17		0.10																	
EMX1_13		G	A	G	T	C	C	a	A	G	-	A	G	A	A	G	A	A	G	t	g	A	G	G	0.00	0.00	Invalidated	Invalidated
	Untreated					0.08	0.12																					
	BE1					0.07	0.11																					
	BE2					0.07	0.11																					
	BE3					0.08	0.13																					
EMX1_14		G	A	G	T	C	C	t	A	G	-	A	G	A	A	G	A	A	G	g	A	A	G	G	0.01	0.01	Invalidated	Invalidated
	Untreated					0.06	0.13																					
	BE1					0.09	0.17																					
	BE2					0.05	0.10																					
	BE3					0.05	0.13																					
EMX1_15		G	A	a	T	C	C	a	A	G	C	A	G	g	A	G	A	A	G	A	A	G	G	A	0.46	0.89	Validated	Validated
	Untreated					0.04	0.07				0.05																	
	BE1					0.03	0.08				0.06																	
	BE2					0.04	0.07				0.06																	
	BE3					0.14	0.18				0.05																	
EMX1_16		G	t	a	c	C	a	G	A	G	-	A	G	A	A	G	A	A	G	A	g	A	G	G	0.00	0.00	Invalidated	Invalidated
	Untreated				0.06	0.06																						
	BE1				0.05	0.05																						
	BE2				0.05	0.05																						
	BE3				0.05	0.05																						
EMX1_17		G	A	G	T	C	C	c	A	G	C	A	a	A	A	G	A	A	G	A	A	A	A	G	0.01	0.00	Invalidated	Invalidated
	Untreated					0.10	0.19	0.09			0.07																	
	BE1					0.13	0.17	0.09			0.05																	
	BE2					0.10	0.20	0.06			0.03																	
	BE3					0.11	0.20	0.07			0.07																	
EMX1_18		a	A	G	T	C	C	a	A	G	t	-	G	A	A	G	A	A	G	A	A	A	G	G	0.01	0.01	Validated	Invalidated
	Untreated					0.05	0.09																					
	BE1					0.08	0.09																					
	BE2					0.08	0.10																					
	BE3					0.09	0.11																					
EMX1_19		a	A	G	T	C	C	a	t	G	C	A	G	A	A	G	A	g	G	A	A	G	G	G	0.01	0.02	Validated	Invalidated
	Untreated					0.03	0.07				0.10																	
	BE1					0.17	0.10				0.12																	
	BE2					0.09	0.14				0.08																	
	BE3					0.24	0.30				0.12																	
EMX1_20		G	A	G	T	C	C	t	A	G	-	A	G	A	A	G	A	A	a	A	A	G	G	G	0.27	0.25	Validated	Invalidated
	Untreated					0.05	0.12																					
	BE1					0.28	0.24																					
	BE2					0.39	0.42																					
	BE3					0.50	0.57																					

(*Continued*)

		G	A	G	T	C	C	c	t	-	C	A	G	g	A	G	A	A	G	A	A	A	G	G				
EMX1_21	Untreated					0.16	0.08	0.07			0.03														0.00	0.00	Invalidated	Invalidated
	BE1					0.15	0.10	0.06			0.04																		
	BE2					0.20	0.13	0.11			0.05																		
	BE3					0.20	0.12	0.10			0.06																		
		a	c	G	T	C	t	G	A	G	C	A	G	G	A	A	G	A	A	G	A	A	T	G	G				
EMX1_22	Untreated		0.14			0.04					0.11														0.02	0.17	Validated	Validated
	BE1		0.17			0.36					0.10																		
	BE2		0.13			0.14					0.11																		
	BE3		0.15			0.62					0.12																		
		G	A	G	T	t	C	c	A	G	a	A	G	A	A	G	A	A	G	A	A	G	A	G				
EMX1_23	Untreated						0.06	0.08																	0.01	0.01	Invalidated	Invalidated
	BE1						0.09	0.13																					
	BE2						0.06	0.10																					
	BE3						0.06	0.09																					
		G	G	A	G	T	C	C	t	A	a	-	A	G	A	A	G	A	A	G	c	A	G	G	G				
EMX1_24	Untreated						0.05	0.18													0.11				0.00	0.00	Invalidated	Invalidated
	BE1						0.04	0.18													0.12								
	BE2						0.05	0.19													0.11								
	BE3						0.05	0.22													0.12								
		c	A	G	T	C	C	a	A	a	C	A	G	A	A	G	A	g	G	A	A	T	G	G				
EMX1_25	Untreated	0.11				0.05	0.11				0.11														1.06	1.04	Validated	Invalidated
	BE1	0.08				0.10	0.10				0.10																		
	BE2	0.10				0.06	0.10				0.11																		
	BE3	0.11				0.07	0.13				0.11																		

FANCF

Base editing efficiency, C to other bases (%)

		G	G	A	A	T	C	C	C	T	T	C	T	G	C	A	G	C	A	C	C	T	G	G	Indel frequency (%) (-) RGEN	Indel frequency (%) RGEN	Validation BE3	Validation Cas9
On-target (FANCF_2)	Untreated						0.06	0.10	0.04			0.03			0.13			0.13		0.05	0.04				0.01	44.48	Validated	Validated
	BE1						0.81	0.39	0.42			0.07			0.12			0.13		0.06	0.03							
	BE2						2.11	2.06	1.97			0.39			0.14			0.09		0.07	0.02							
	BE3						10.26	9.44	9.28			4.12			0.18			0.12		0.05	0.04							
		t	G	A	A	T	C	C	C	a	T	C	T	c	C	A	G	C	A	C	C	A	G	G				
FNACF_1	Untreated						0.07	0.10	0.08			0.03		0.04	0.04			0.07		0.04	0.07				0.00	0.02	Validated	Validated
	BE1						0.07	0.10	0.09			0.03		0.05	0.05			0.09		0.03	0.06							
	BE2						0.10	0.10	0.12			0.03		0.02	0.06			0.07		0.02	0.08							
	BE3						0.16	0.16	0.18			0.06		0.05	0.07			0.09		0.03	0.07							
		G	G	A	g	T	C	C	C	T	c	C	T	a	C	A	G	C	A	C	C	A	G	G				
FNACF_3	Untreated						0.06	0.09	0.05		0.08	0.03			0.06			0.07		0.06	0.14				0.01	0.37	Validated	Validated
	BE1						0.06	0.09	0.06		0.08	0.03			0.06			0.06		0.07	0.13							
	BE2						0.06	0.10	0.05		0.08	0.04			0.05			0.07		0.07	0.15							
	BE3						0.20	0.23	0.18		0.16	0.06			0.08			0.06		0.07	0.18							
		G	G	A	g	T	C	C	C	T	c	C	T	a	C	A	G	C	A	C	C	A	G	G				
FNACF_4	Untreated						0.06	0.05	0.05		0.05	0.06			0.06			0.03		0.03	0.06				0.01	0.22	Validated	Validated
	BE1						0.06	0.05	0.05		0.02	0.07			0.07			0.03		0.03	0.05							
	BE2						0.07	0.06	0.03		0.05	0.06			0.07			0.02		0.04	0.06							
	BE3						0.11	0.09	0.12		0.05	0.06			0.07			0.04		0.04	0.04							
		G	G	A	A	T	C	C	C	T	T	C	T	a	C	A	G	C	A	t	C	C	T	G				
FNACF_5	Untreated						0.09	0.07	0.05			0.03			0.07			0.03			0.03				0.00	0.00	Invalidated	Invalidated
	BE1						0.07	0.06	0.04			0.03			0.07			0.03			0.03							
	BE2						0.08	0.05	0.06			0.03			0.06			0.05			0.03							
	BE3						0.10	0.07	0.05			0.03			0.07			0.03			0.02							
		G	G	A	g	T	C	C	C	T	c	C	T	G	C	A	G	C	A	C	C	T	G	A				
FNACF_6	Untreated						0.04	0.04	0.04		0.02	0.04			0.09			0.06		0.02	0.04				0.00	0.28	Validated	Validated
	BE1						0.05	0.05	0.02		0.02	0.04			0.12			0.04		0.05	0.05							
	BE2						0.04	0.05	0.05		0.03	0.06			0.11			0.06		0.05	0.05							
	BE3						0.13	0.09	0.09		0.05	0.06			0.12			0.06		0.05	0.03							
		G	G	A	A	e	C	C	C	g	T	C	T	G	C	A	G	C	A	C	C	A	G	G				
FNACF_7	Untreated					0.03	0.07	0.07	0.06			0.03			0.20			0.05		0.03	0.07				0.01	12.06	Validated	Validated
	BE1					0.05	0.06	0.04	0.07			0.01			0.21			0.05		0.02	0.05							
	BE2					0.04	0.08	0.05	0.08			0.02			0.23			0.06		0.02	0.05							
	BE3					1.06	1.07	1.07	1.02			0.71			0.22			0.07		0.03	0.07							
		G	t	c	t	c	C	C	C	T	T	C	T	G	C	A	G	C	A	C	C	A	G	G				
FNACF_8	Untreated			0.02		0.03	0.04	0.02	0.05			0.03			0.11			0.03		0.02	0.03				0.03	0.05	Invalidated	Invalidated
	BE1			0.02		0.02	0.03	0.04	0.05			0.03			0.08			0.04		0.02	0.04							
	BE2			0.01		0.02	0.02	0.05	0.05			0.02			0.09			0.03		0.02	0.04							
	BE3			0.02		0.02	0.02	0.04	0.08			0.03			0.10			0.03		0.02	0.03							
		a	a	A	A	T	C	C	C	T	T	C	c	G	C	A	G	C	A	C	C	T	A	G				
FNACF_9	Untreated						0.07	0.02	0.04			0.05	0.06		0.05			0.04		0.05	0.06				0.00	0.08	Validated	Validated
	BE1						0.08	0.03	0.04			0.04	0.07		0.04			0.04		0.05	0.04							
	BE2						0.08	0.02	0.03			0.03	0.07		0.04			0.04		0.04	0.06							
	BE3						0.10	0.04	0.05			0.05	0.06		0.06			0.04		0.05	0.03							
		t	G	t	A	T	t	t	C	T	T	C	T	G	C	c	t	C	A	g	g	C	T	G				
FNACF_10	Untreated																											
	BE1																											
	BE2																											
	BE3																											
		G	G	A	A	T	a	t	C	T	T	C	T	G	C	A	G	C	c	C	C	A	G	G				
FNACF_11	Untreated								0.03			0.05			0.22			0.03	0.05	0.05	0.10				0.02	0.03	Invalidated	Invalidated
	BE1								0.03			0.04			0.23			0.03	0.06	0.05	0.09							
	BE2								0.03			0.04			0.21			0.03	0.06	0.07	0.09							
	BE3								0.04			0.03			0.21			0.02	0.06	0.05	0.09							
		G	a	g	t	g	C	C	C	T	g	a	a	G	C	c	t	C	A	g	C	T	G	G				
FNACF_12	Untreated																											
	BE1																											
	BE2																											
	BE3																											
		a	c	c	A	T	C	C	C	T	c	C	T	G	C	A	G	C	A	C	C	A	G	G				
FNACF_13	Untreated		0.07	0.06			0.04	0.04	0.04		0.06	0.05			0.10			0.03		0.08	0.04				0.01	0.03	Validated	Validated
	BE1		0.14	0.07			0.07	0.04	0.05		0.06	0.04			0.10			0.04		0.06	0.05							
	BE2		0.11	0.07			0.04	0.03	0.04		0.06	0.04			0.12			0.04		0.05	0.05							
	BE3		0.13	0.08			0.15	0.15	0.14		0.13	0.09			0.10			0.04		0.06	0.04							
		t	G	A	A	T	C	C	t	a	a	C	T	G	C	A	G	C	A	C	C	A	G	G				
FNACF_14	Untreated						0.09	0.05				0.04			0.09			0.06		0.08	0.06				0.00	0.00	Invalidated	Invalidated
	BE1						0.09	0.04				0.05			0.10			0.06		0.10	0.07							
	BE2						0.07	0.05				0.04			0.07			0.06		0.09	0.06							
	BE3						0.10	0.08				0.03			0.11			0.07		0.10	0.07							
		c	t	c	t	g	t	C	C	T	T	C	T	G	C	A	G	C	A	C	C	T	G	G				
FNACF_15	Untreated	0.03		0.04				0.05	0.02			0.02			0.06			0.02		0.01	0.03				0.02	0.00	Invalidated	Invalidated
	BE1	0.03		0.02				0.04	0.02			0.02			0.06			0.03		0.02	0.04							
	BE2	0.04		0.03				0.04	0.03			0.02			0.05			0.03		0.02	0.04							
	BE3	0.03		0.02				0.07	0.03			0.02			0.05			0.03		0.02	0.04							

(*Continued*)

HBB

		Base editing efficiency, C to other bases (%)																						Indel frequency (%)		Validation	
																								(-) RGEN	RGEN	BE3	Cas9
On-target (HBB_1)		C	T	T	G	C	C	C	C	A	C	A	G	G	G	C	A	G	T	A	A	C	G	G			
	Untreated	0.05				0.08	0.03	0.05	0.04		0.04					0.08											
	BE1	0.04				0.08	0.14	0.17	0.08		0.05					0.07											
	BE2	0.08				0.56	0.80	0.83	0.80		0.07					0.06									0.02	38.35	Validated	Validated
	BE3	0.10				3.01	4.51	4.88	4.64		0.14					0.08											
HBB_2		t	T	g	c	t	C	C	C	A	C	A	G	G	G	C	A	G	T	A	A	A	C	G			
	Untreated				0.07		0.06	0.04	0.05		0.04					0.06											
	BE1				0.07		0.09	0.07	0.07		0.04					0.07											
	BE2				0.14		0.24	0.22	0.22		0.05					0.06									0.02	0.01	Validated	Invalidated
	BE3				0.42		0.89	0.84	0.86		0.07					0.06											
HBB_3		g	c	T	G	C	C	C	C	A	C	A	G	G	G	C	A	G	c	A	A	A	G	G			
	Untreated		0.07			0.06	0.06	0.11	0.03		0.07					0.14			0.09							
	BE1		0.08			0.06	0.06	0.10	0.03		0.05					0.10			0.08							
	BE2		0.09			0.13	0.15	0.17	0.09		0.05					0.12			0.09					0.01	3.57	Validated	Validated
	BE3		0.09			0.80	0.86	0.87	0.75		0.07					0.11			0.09							
HBB_4		g	T	g	G	C	C	C	C	A	C	A	G	G	G	C	A	G	g	A	A	T	G	G			
	Untreated					0.07	0.13	0.06	0.09		0.04					0.06											
	BE1					0.09	0.14	0.07	0.08		0.05					0.08											
	BE2					0.09	0.15	0.08	0.12		0.04					0.07									0.00	0.70	Validated	Validated
	BE3					0.14	0.20	0.13	0.16		0.07					0.08											
HBB_5		a	T	T	G	C	C	C	C	A	C	g	G	G	G	C	A	G	T	g	A	C	G	G			
	Untreated					0.12	0.19	0.73	0.40		0.16					0.20											
	BE1					0.16	0.20	0.76	0.47		0.19					0.25											
	BE2					0.14	0.16	0.77	0.51		0.17					0.28									0.00	0.35	Validated	Validated
	BE3					0.36	0.42	0.95	0.73		0.20					0.21											
HBB_6		a	c	T	c	t	C	C	C	A	C	A	a	G	G	C	A	G	T	A	A	G	G	G			
	Untreated		0.11		0.12	0.08	0.11	0.20	0.08		0.05					0.17											
	BE1		0.10		0.16	0.10	0.09	0.20	0.10		0.04					0.14											
	BE2		0.08		0.16	0.11	0.11	0.21	0.10		0.05					0.20									0.02	0.01	Invalidated	Invalidated
	BE3		0.10		0.14	0.13	0.13	0.22	0.09		0.05					0.17											
HBB_7		t	c	a	G	C	C	C	C	A	C	A	G	G	G	C	A	G	T	A	A	G	G	G			
	Untreated		0.03			0.07	0.07	0.09	0.05		0.05					0.08											
	BE1		0.14			0.09	0.09	0.11	0.06		0.06					0.14											
	BE2		0.27			0.09	0.22	0.25	0.19		0.05					0.09									0.00	20.92	Validated	Validated
	BE3		2.82			0.80	2.89	4.01	4.20		0.14					0.09											

HEK2

		Base editing efficiency, C to other bases (%)																						Indel frequency (%)		Validation	
																								(-) RGEN	RGEN	BE3	Cas9
On-target (HEK2_2)		G	A	A	C	A	C	A	A	A	G	C	A	T	A	G	A	C	T	G	C	G	G	G			
	Untreated				0.04		0.05					0.04						0.16									
	BE1				0.65		10.29					0.04						0.18									
	BE2				7.32		14.69					0.03						0.17							0.00	43.28	Validated	Validated
	BE3				11.74		22.29					0.07						0.18									
HEK2_1		G	A	A	C	A	C	A	A	t	G	C	A	T	A	G	A	t	T	G	C	C	G	G			
	Untreated				0.10		0.09					0.11									0.18						
	BE1				0.10		0.10					0.13									0.21				0.00	1.01	Validated	Validated
	BE2				0.13		0.12					0.11									0.16						
	BE3				0.17		0.21					0.11									0.19						
HEK2_3		a	A	c	t	c	C	A	A	A	G	C	A	T	A	t	A	C	T	G	C	T	G	G			
	Untreated			0.09		0.09	0.34					0.25									0.09						
	BE1			0.08		0.07	0.37					0.24									0.08						
	BE2			0.09		0.07	0.38					0.19									0.08				0.00	0.00	Invalidated	Invalidated
	BE3			0.09		0.07	0.38					0.24									0.07						

HEK3

		Base editing efficiency, C to other bases (%)																						Indel frequency (%)		Validation	
																								(-) RGEN	RGEN	BE3	Cas9
On-target (HEK3_2)		G	G	C	C	C	A	G	A	C	T	G	A	G	C	A	C	G	T	G	A	T	G	G			
	Untreated			0.13	0.46	0.42				0.14					0.10		0.07										
	BE1			0.38	6.45	8.56				0.59					0.14		0.08										
	BE2			0.37	6.27	8.17				0.41					0.20		0.06								0.00	60.16	Validated	Validated
	BE3			1.00	24.71	31.39				0.76					0.09		0.10										
HEK3_1		a	G	C	t	C	A	G	A	C	T	G	A	G	C	A	a	G	T	G	A	G	G	G			
	Untreated			0.12		0.04				0.04					0.14												
	BE1			0.12		0.04				0.07					0.13												
	BE2			0.13		0.05				0.08					0.17										0.00	2.93	Invalidated	Validated
	BE3			0.13		0.09				0.05					0.13												
HEK3_3		G	t	g	g	C	c	c	A	g	a	G	A	G	C	A	C	G	T	G	t	G	G	G			
	Untreated					0.07	0.06	0.07							0.12		0.13										
	BE1					0.08	0.05	0.10							0.09		0.11										
	BE2					0.08	0.05	0.06							0.11		0.12								0.00	0.00	Invalidated	Invalidated
	BE3					0.07	0.05	0.07							0.10		0.10										
HEK3_4		c	a	C	C	C	A	G	A	C	T	G	A	G	C	A	C	G	T	G	c	T	G	G			
	Untreated	0.08		0.07	0.07	0.05				0.01					0.14		0.06				0.04						
	BE1	0.09		0.06	0.08	0.06				0.03					0.13		0.04				0.04						
	BE2	0.09		0.07	0.07	0.06				0.02					0.10		0.05				0.05			0.00	4.16	Invalidated	Validated
	BE3	0.08		0.05	0.08	0.06				0.02					0.13		0.05				0.05						
HEK3_5		c	G	g	C	C	c	a	A	C	T	G	A	G	C	A	a	G	T	G	A	T	G	G			
	Untreated	0.16			0.08	0.13	0.10			0.06					0.19												
	BE1	0.19			0.11	0.14	0.07			0.06					0.21										0.00	0.00	Invalidated	Invalidated
	BE2	0.16			0.08	0.13	0.09			0.05					0.16												
	BE3	0.16			0.08	0.13	0.09			0.05					0.20												
HEK3_6		a	G	a	C	C	A	G	A	C	T	G	A	G	C	A	a	G	a	G	A	G	G	G			
	Untreated				0.08	0.10				0.06					0.20												
	BE1				0.09	0.12				0.06					0.19										0.00	0.02	Invalidated	Invalidated
	BE2				0.08	0.12				0.06					0.19												
	BE3				0.10	0.11				0.05					0.16												
HEK3_7		G	G	C	C	a	c	t	c	a	T	G	g	c	C	A	C	a	T	a	c	T	G	G			
	Untreated			0.45	0.15		0.05		0.19					0.29	0.26						0.06						
	BE1			0.45	0.16		0.08		0.19					0.30	0.28						0.06				0.00	0.00	Invalidated	Invalidated
	BE2			0.47	0.17		0.09		0.19					0.31	0.24						0.06						
	BE3			0.44	0.16		0.08		0.19					0.29	0.26						0.06						

(*Continued*)

HEK4

		Base editing efficiency, C to other bases (%)																						Indel frequency (%)		Validation	
		G	G	C	A	C	T	G	C	G	G	C	T	G	G	A	G	G	T	G	G	G	G	(-) RGEN	RGEN	BE3	Cas9
On-target (HEK4_1)	Untreated			0.16		0.11			0.20			0.07												0.00	59.38	Validated	Validated
	BE1			0.17		6.18			0.25			0.07															
	BE2			0.65		10.35			0.84			0.06															
	BE3			2.34		41.18			0.80			0.07															
		G	G	C	A	C	T	G	C	t	G	C	T	G	G	g	G	G	T	G	G	T	G				
HEK4_2	Untreated			0.11		0.05			0.15			0.98												0.02	35.65	Validated	Validated
	BE1			0.13		0.38			0.14			0.98															
	BE2			0.16		0.46			0.13			0.93															
	BE3			0.31		5.93			0.22			1.07															
		G	G	C	A	C	T	G	C	a	-	C	T	G	G	A	G	G	T	t	G	T	G				
HEK4_3	Untreated			0.08		0.05			0.07			0.05												0.00	0.00	Validated	Invalidated
	BE1			0.10		0.22			0.09			0.05															
	BE2			0.11		0.22			0.07			0.05															
	BE3			0.09		0.39			0.08			0.03															
		G	G	C	t	C	T	G	C	G	G	C	T	G	G	A	G	G	g	G	G	T	G				
HEK4_4	Untreated			0.04		0.05			0.34			0.13												0.07	29.61	Validated	Validated
	BE1			0.05		0.26			0.35			0.13															
	BE2			0.06		0.19			0.35			0.15															
	BE3			0.07		2.07			0.34			0.17															
		a	G	C	A	C	T	G	C	a	G	a	T	G	G	A	G	G	a	G	G	C	G				
HEK4_5	Untreated			0.08		0.07			0.11															0.00	0.08	Validated	Validated
	BE1			0.09		0.11			0.11																		
	BE2			0.09		0.07			0.10																		
	BE3			0.10		0.52			0.20																		
		G	G	C	A	C	T	G	C	G	G	C	a	G	G	g	a	G	g	a	G	G	G				
HEK4_6	Untreated																										
	BE1																										
	BE2																										
	BE3																										
		t	G	C	A	C	T	G	C	G	G	C	c	G	G	A	G	G	a	G	G	T	G				
HEK4_7	Untreated			0.21		0.12			0.36			0.14	0.09											0.02	35.87	Validated	Validated
	BE1			0.15		0.53			0.31			0.13	0.08														
	BE2			0.19		1.25			0.32			0.11	0.14														
	BE3			0.37		10.75			0.41			0.12	0.07														
		G	G	C	A	C	T	-	g	G	G	C	T	G	a	A	G	G	T	a	G	A	G				
HEK4_8	Untreated			0.09		0.05						0.08												0.04	0.04	Validated	Invalidated
	BE1			0.07		0.15						0.05															
	BE2			0.08		0.17						0.07															
	BE3			0.07		0.18						0.06															
		G	G	C	A	C	T	G	t	G	G	C	T	G	c	A	G	G	T	G	G	A	G				
HEK4_9	Untreated			0.09		0.03						0.02			0.04									0.02	25.09	Invalidated	Validated
	BE1			0.08		0.04						0.04			0.03												
	BE2			0.12		0.03						0.04			0.03												
	BE3			0.12		0.02						0.04			0.05												
		t	G	C	t	C	T	G	C	G	G	C	a	G	G	A	G	G	a	G	G	A	G				
HEK4_10	Untreated			0.08		0.17			0.06			0.06												2.67	3.08	Invalidated	Validated
	BE1			0.07		0.17			0.06			0.05															
	BE2			0.08		0.18			0.07			0.07															
	BE3			0.08		0.19			0.07			0.07															
		a	G	C	A	C	T	G	C	a	G	C	T	G	G	g	a	G	T	G	G	A	G				
HEK4_11	Untreated			0.16		0.05			0.13			0.07												0.04	8.97	Validated	Validated
	BE1			0.12		0.47			0.12			0.07															
	BE2			0.13		0.64			0.14			0.08															
	BE3			0.19		1.83			0.18			0.08															
		G	G	C	A	C	T	G	a	G	G	g	T	G	G	A	G	G	T	G	G	G	G				
HEK4_12	Untreated			0.10		0.03																		0.08	10.38	Validated	Validated
	BE1			0.07		0.65																					
	BE2			0.10		0.47																					
	BE3			0.09		0.99																					
		G	G	C	A	C	T	G	g	G	G	C	T	G	G	A	G	a	c	G	G	G	G				
HEK4_13	Untreated			0.13		0.15						0.13							0.23					0.11	0.69	Invalidated	Validated
	BE1			0.12		0.14						0.11							0.18								
	BE2			0.10		0.13						0.09							0.15								
	BE3			0.11		0.12						0.12							0.18								
		a	G	g	A	C	T	G	C	G	G	C	T	G	G	g	G	G	T	G	G	T	G				
HEK4_14	Untreated					0.06			0.28			0.05												0.38	46.26	Validated	Validated
	BE1					0.50			0.37			0.03															
	BE2					0.63			0.38			0.04															
	BE3					5.20			0.50			0.04															
		G	G	C	A	C	T	G	C	a	a	C	T	G	G	A	a	G	T	G	a	T	G				
HEK4_15	Untreated			0.11		0.06			0.12			0.03												0.01	0.14	Validated	Validated
	BE1			0.10		0.08			0.07			0.02															
	BE2			0.08		0.08			0.08			0.02															
	BE3			0.10		0.26			0.09			0.03															
		G	G	C	A	C	T	G	g	G	G	t	T	G	G	A	G	G	T	G	G	G	G				
HEK4_16	Untreated			0.17		0.16																		0.12	25.87	Validated	Validated
	BE1			0.14		1.01																					
	BE2			0.17		0.58																					
	BE3			0.38		3.41																					
		G	c	C	A	C	T	G	C	a	G	C	T	a	G	A	G	G	T	G	G	A	G				
HEK4_17	Untreated		0.14	0.05		0.07			0.20			0.03												0.01	2.93	Validated	Validated
	BE1		0.10	0.06		0.24			0.13			0.04															
	BE2		0.09	0.10		0.27			0.14			0.04															
	BE3		0.12	0.34		3.12			0.22			0.04															
		G	c	C	A	C	T	G	C	G	a	C	T	G	G	A	G	G	a	G	G	G	G				
HEK4_18	Untreated		0.14	0.07		0.06			60.77			0.04												0.16	0.37	Validated	Validated
	BE1		0.10	0.05		0.08			61.73			0.03															
	BE2		0.12	0.03		0.05			60.63			0.05															
	BE3		0.10	0.08		0.12			60.98			0.04															
		G	G	C	A	C	T	G	-	G	G	C	T	G	G	A	G	G	c	G	G	G	G				
HEK4_19	Untreated			0.06		0.06						0.05							0.12					0.10	0.11	Invalidated	Invalidated
	BE1			0.07		0.04						0.06							0.11								
	BE2			0.08		0.06						0.04							0.10								
	BE3			0.08		0.05						0.07							0.09								
		a	G	C	t	C	T	G	C	G	G	C	a	G	G	A	G	t	T	G	G	A	G				
HEK4_20	Untreated			0.24		0.02			0.20			0.12												0.02	0.07	Invalidated	Validated
	BE1			0.21		0.03			0.20			0.08															
	BE2			0.21		0.02			0.17			0.08															
	BE3			0.23		0.02			0.22			0.11															

RNF2

		Base editing efficiency (%)																						Indel frequency (%)		Validation	
		G	T	C	A	T	C	T	T	A	G	T	C	A	T	T	A	C	C	T	G	A	G	G	(-) RGEN	RGEN	BE3	Cas9
On-target (RNF2_1)	Untreated						0.07						0.06					0.03	0.07						0.03	66.13	Validated	Validated
	BE1						2.90						0.08					0.03	0.07									
	BE2						3.89						0.62					0.05	0.08									
	BE3						31.12						3.45					0.16	0.08									

Table 3. Base editing efficiencies at Digenome-captured sites associated only with 3 different Cas9 nucleases. No substitutions were detectably induced by BE3 at these Cas9-associated sites. On-target sequences (*EXM1*_On, *HBB*_On, and *RNF2*_ON) are also shown.

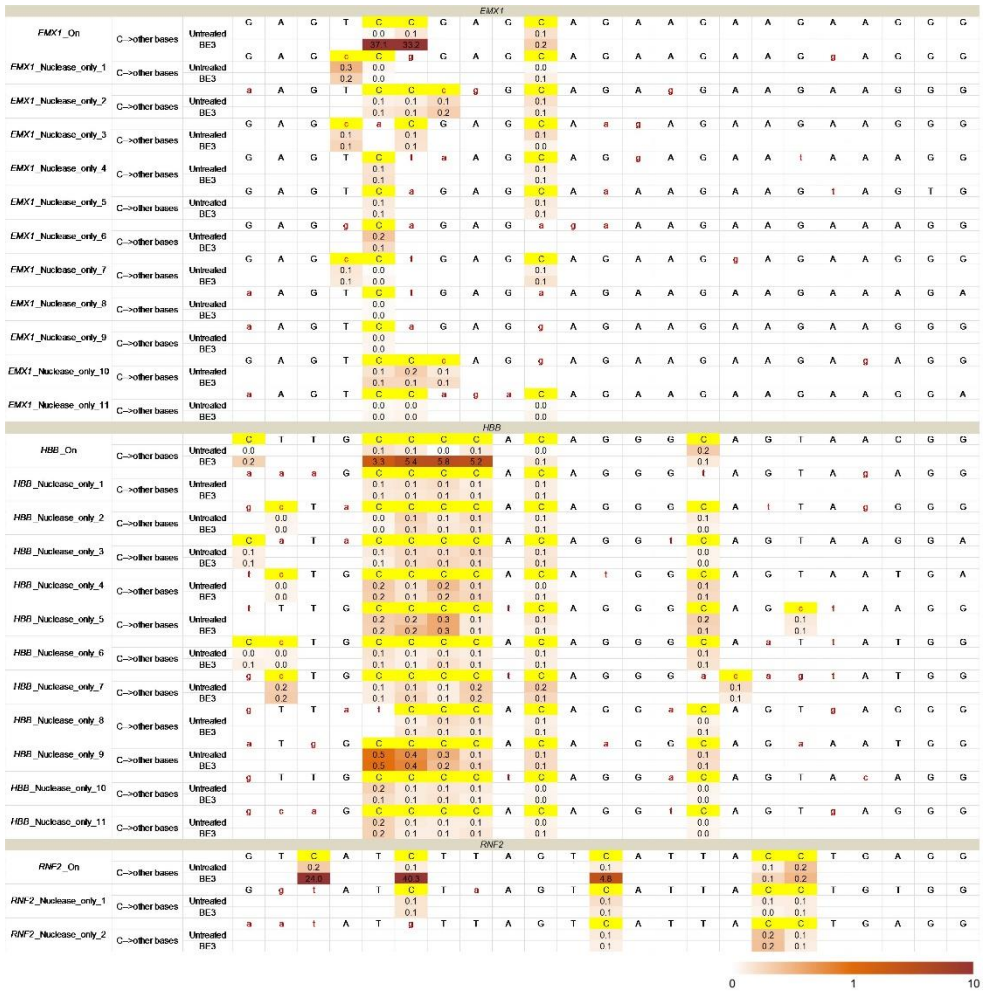
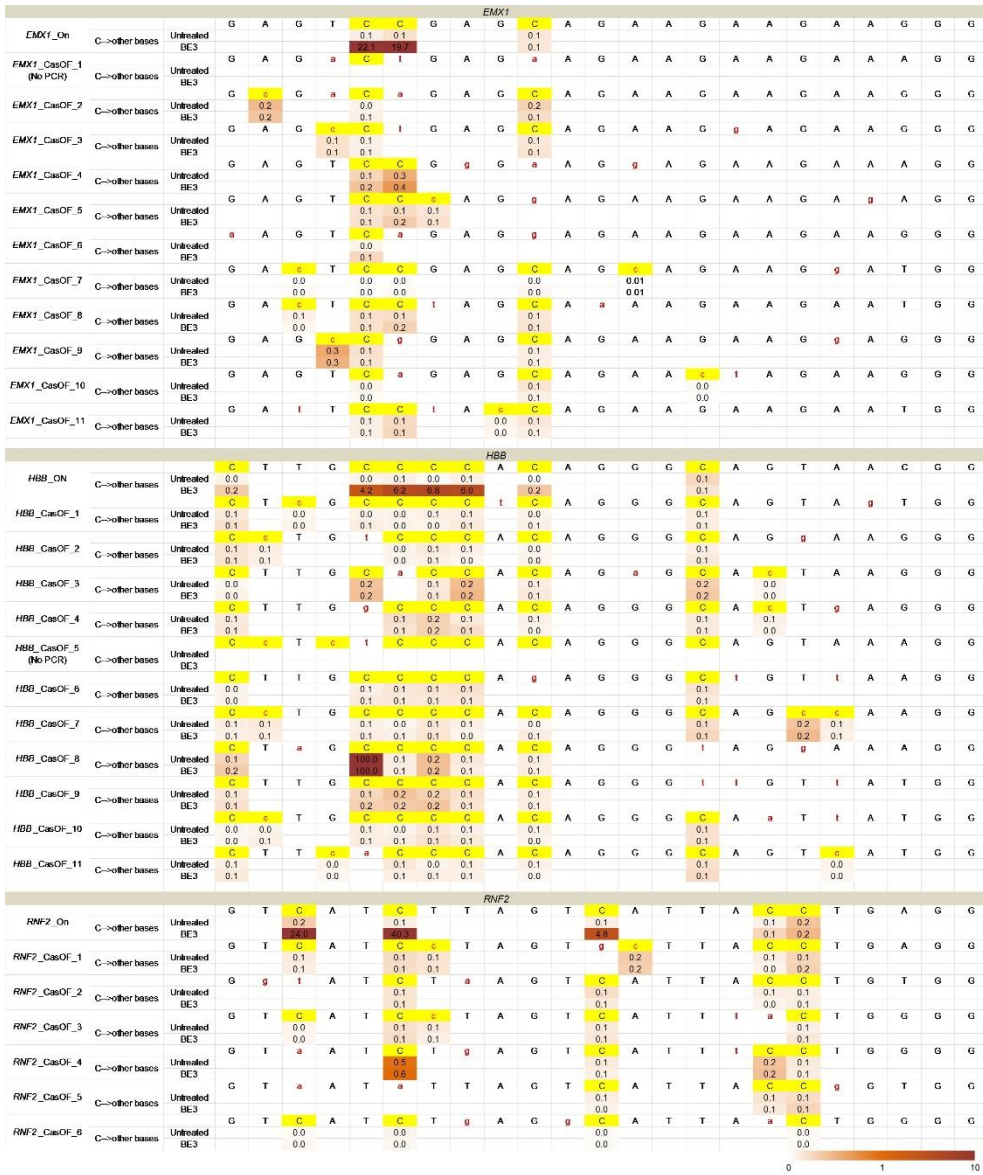


Table 4. Base editing efficiencies of 6 different BE3 deaminases at Digenome-negative sites with ≤ 3 mismatches with respective on-target sequences. No substitutions were detectably induced by BE3 at these sites. On-target sequences are also shown.



(*Continued*)

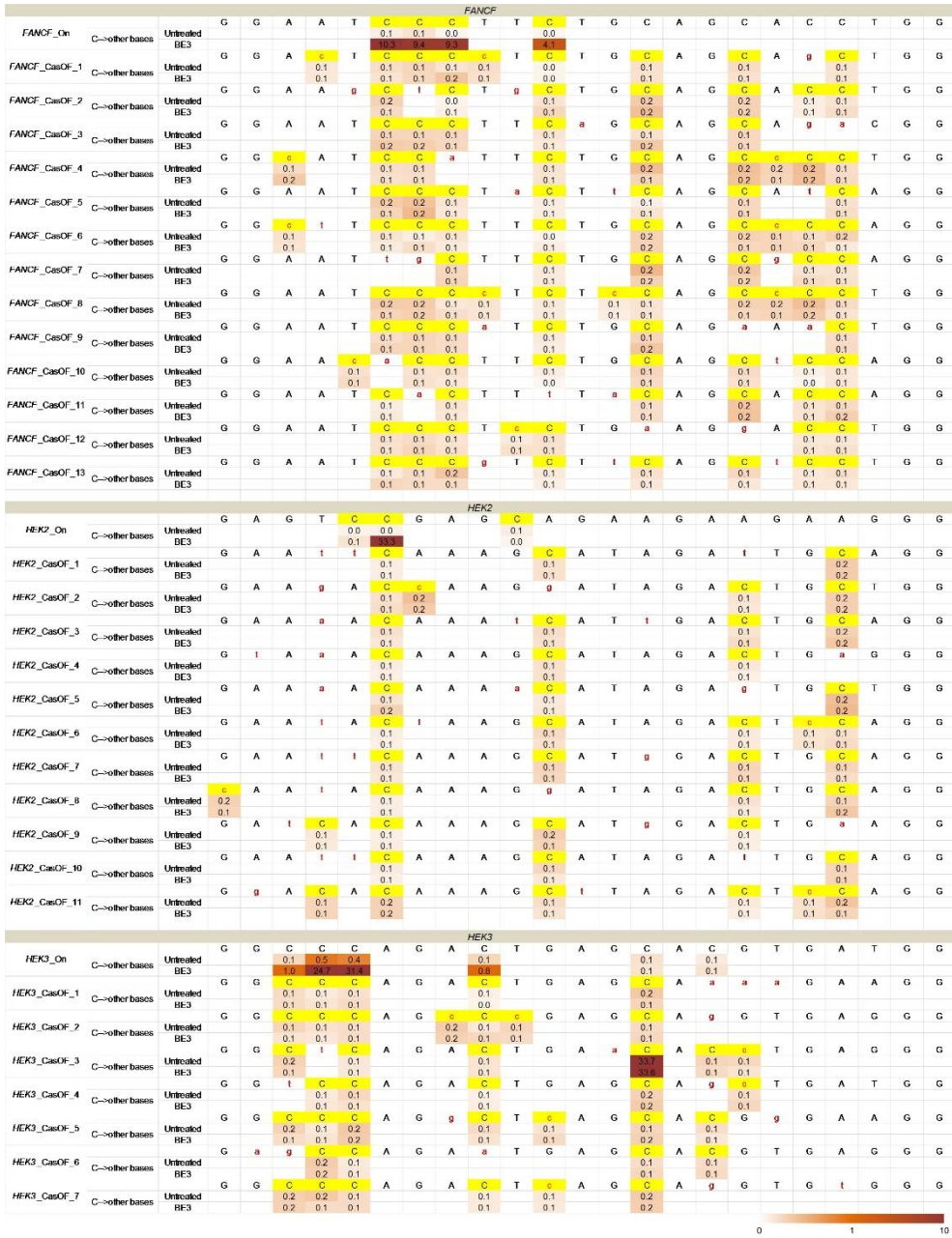
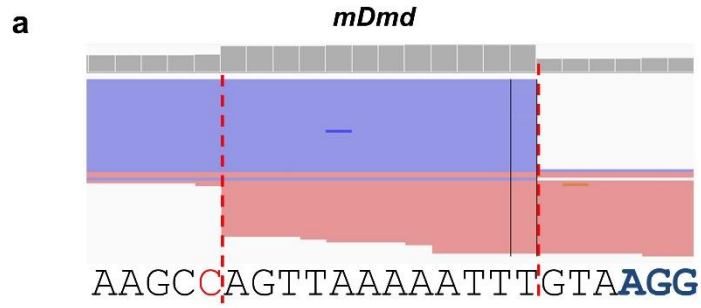

FANCF
FANCF_On
FANCF_CasOF_1
FANCF_CasOF_2
FANCF_CasOF_3
FANCF_CasOF_4
FANCF_CasOF_5
FANCF_CasOF_6
FANCF_CasOF_7
FANCF_CasOF_8
FANCF_CasOF_9
FANCF_CasOF_10
FANCF_CasOF_11
FANCF_CasOF_12
FANCF_CasOF_13
C-->other bases
Untreated
BE3
HEK2
HEK2_On
HEK2_CasOF_1
HEK2_CasOF_2
HEK2_CasOF_3
HEK2_CasOF_4
HEK2_CasOF_5
HEK2_CasOF_6
HEK2_CasOF_7
HEK2_CasOF_8
HEK2_CasOF_9
HEK2_CasOF_10
HEK2_CasOF_11
HEK3
HEK3_On
HEK3_CasOF_1
HEK3_CasOF_2
HEK3_CasOF_3
HEK3_CasOF_4
HEK3_CasOF_5
HEK3_CasOF_6
HEK3_CasOF_7
0
1
10

Table 5. Off-target effect index (OTI)

	Cas9						BE3					
Gene ID	No. of Digenome-positive sites	No of NGS-tested sites	No. of validated sites	Validation rates (%)	OTI	Refs	No. of Digenome-positive sites	No of NGS-tested sites	No. of validated sites	Validation rates (%)	OTI	Refs
EMX1	143	28	12	N.A.	> 1.7	14	25	25	17	68	0.58	this study
FANCF	46	16	9	N.A.	> 0.75	14	13	13	8	62	0.19	this study
HEK2	35	18	3	N.A.	> 0.28	14	3	3	2	67	0	this study
HEK3	31	20	8	N.A.	> 0.76	14	7	7	2	29	0.0053	this study
HEK4	216	41	20	N.A.	> 1.1	14	67	19	14	N.A.	> 0.70	this study
RNF2	13	13	3	23	0.0049	14	1	1	1	100	0	this study
HBB	78	78	5	6.4	2.8	13	7	7	5	71	1.7	this study
VEGFA T1	86	86	9	10	4.4	13			N.A.			
Avarage	81 ± 26			13 ± 6.1			18 ± 9.5			66 ± 10		

(By Desik Kim in Institute for Basic Science)

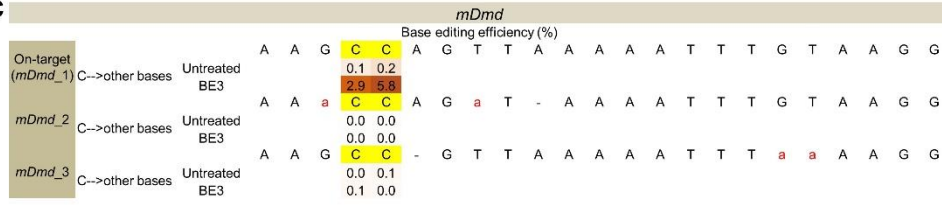


b

mDmd

	Chr	Position	DNA seq at a cleavage sites	DNA Cleavage Score	Bulge
mDmd_1	chrX	83781754	AAGCCAGTTAAAAATTTGTAAGG	10.9	x
mDmd_2	chr14	81071267	AAaCCAGaT-AAAATTTGTAAGG	1.7	RNA bulge
mDmd_3	chr4	100722948	AAGCC-GTTAAAAATTTaaAAGG	0.7	RNA bulge

c



(By Kyoungmi Kim, Seuk-Min Ryu & Desik Kim in Institute for Basic Science)

Figure 18. Digenome-seq to identify off-target site of BE3 in the mouse genome.

(a) IGV image showing straight alignments of sequence reads at the *Dmd* on-target site. (b) Three sites, including the on-target site, were identified by Digenome 2.0. (c) No off-target substitutions were detectably induced at the two candidate sites identified by Digenome-seq in NIH3T3 cells.

4. Reducing BE3 off-target effects via modified sgRNAs

To reduce BE3 off-target effects, I replaced standard GX19 (or gX19) sgRNAs ("g" and "G" represent, respectively, a mismatched and matched guanine) with truncated sgRNAs (Fu et al., 2014) (termed gX18 or gX17) or extended sgRNAs (Cho et al., 2014; Kim et al., 2015; Kim et al., 2016b) containing one or two extra guanines at the 5′ terminus (termed gX20 or ggX20) and measured on-target and off-target base-editing frequencies in HEK293T cells (Figure 19, Figure 20, and Table 6). Truncated sgRNAs reduced off-target effects at many sites but exacerbated them at sites with mismatches at the 5′ terminus (shown by asterisks in Figure 19b and Figure 20b). Extended sgRNAs reduced off-target effects at almost every site without sacrificing on-target effects. Notably, some extended sgRNAs were more active at on-target sites than GX19 (or gX19) sgRNAs (Table 6). Use of attenuated Cas9 variants (Kleinstiver et al., 2016; Rees et al., 2017; Slaymaker et al., 2016) or delivery of BE3 RNPs rather than plasmids (Kim et al., 2016a; Kim et al., 2014; Rees et al., 2017) may further improve the genome-wide specificity of base editing. Digenome-seq, at least in its current form, cannot be used for profiling specificities of BE1 and BE2, which are composed of dCas9 rather than nCas9. Note, however, that these dCas9-based deaminases are substantially less efficient than BE3. Because Digenome-seq is an *in vitro* method, it cannot identify DSB hotspots or uracil hotspots, if any, that may occur in cells independent of Cas9-sgRNA-DNA interactions. New methods may be developed in the future to address these issues.

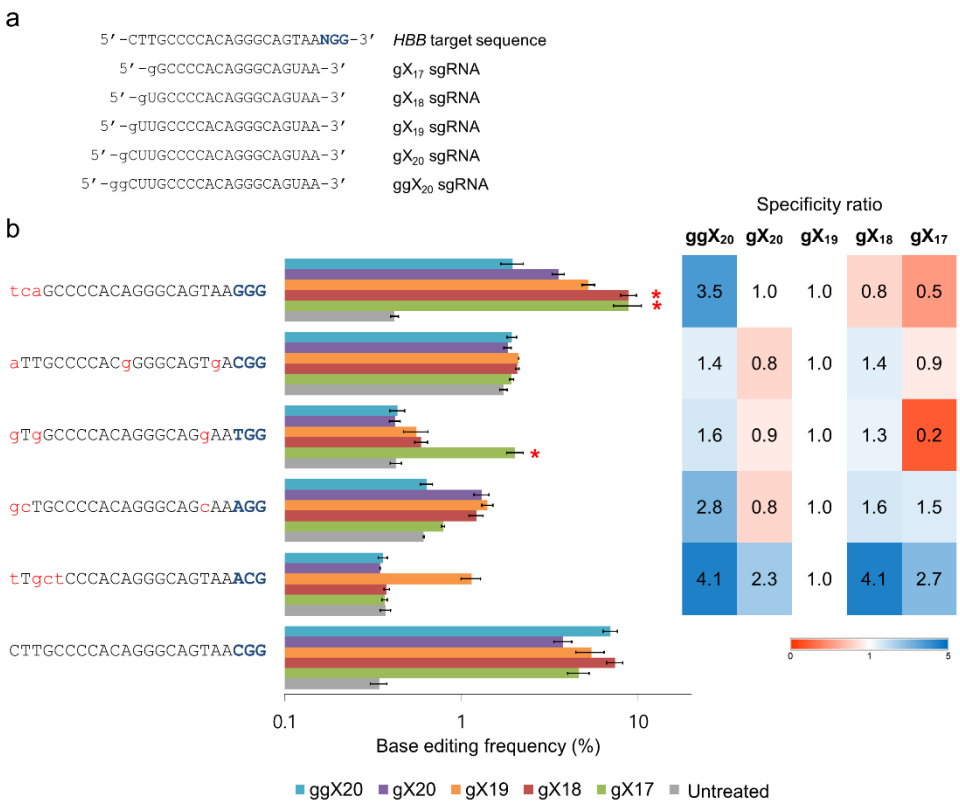


Figure 19. Reducing BE3 off-target effects via modifying length of *HBB*-targeting sgRNAs. (a) Sequences of sgRNAs at the 5′ terminus. (b) Base-editing frequencies at the *HBB* on- and off-target sites in HEK293T cells were measured by targeted deep sequencing. The heatmap represents relative specificities of modified sgRNAs, compared to that of gX19 sgRNA. The specificity ratio was calculated by dividing (on-target frequency of modified sgRNA/off-target frequency of modified sgRNA) by (on-target frequency of gX19 sgRNA/off-target frequency of gX19 sgRNA).

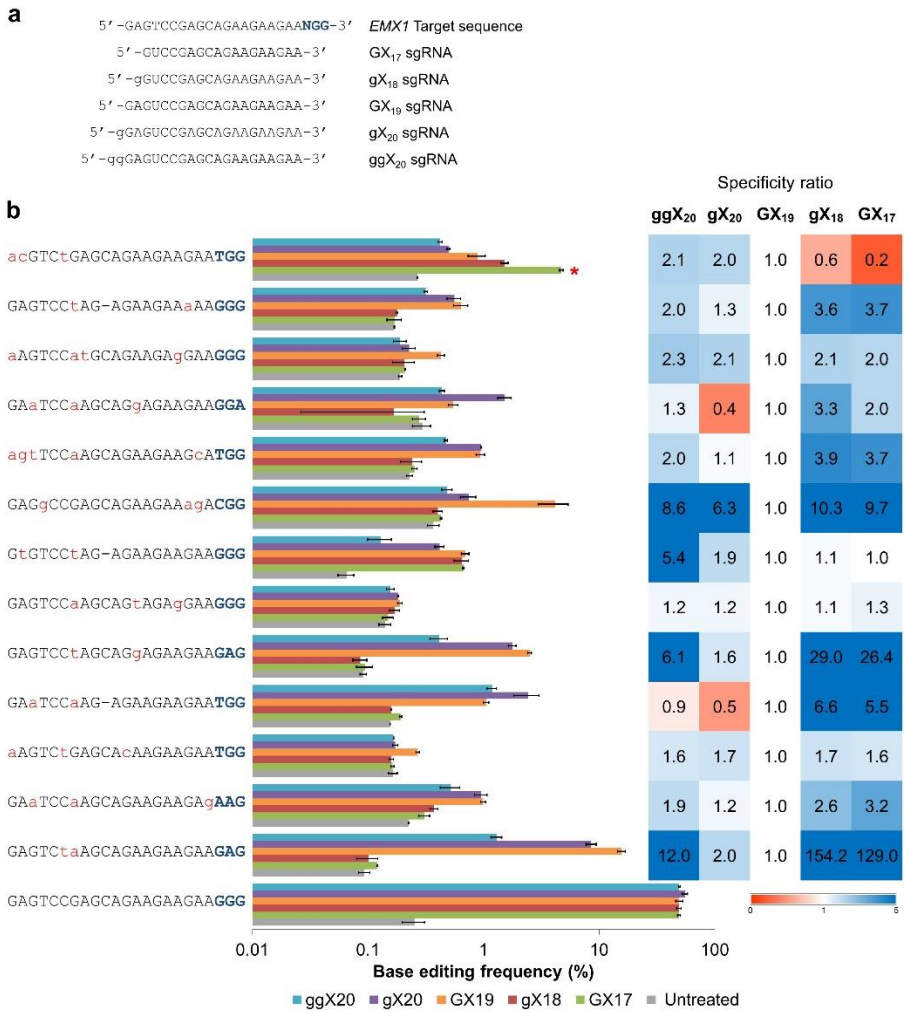


Figure 20. Reducing BE3 off-target effects via modifying length of *EMX1*-targeting sgRNAs. (a) Sequences of sgRNAs at the 5' terminus. (b) Base editing efficiencies were measured at the *EMX1* on- and off-target sites by targeted deep sequencing in HEK293T cells. The heatmap represents relative specificities of modified sgRNAs, compared to that of gX19 sgRNA. The specificity ratio was calculated by dividing (on-target frequency of modified sgRNA/off-target frequency of modified sgRNA) by (on-target frequency of gX19 sgRNA/off-target frequency of gX19 sgRNA).

Table 6. Analysis of BE3 off-target effects via modified sgRNAs.

			EMX1																						
			G	A	G	T	C	C	G	A	G	C	A	G	A	A	G	A	A	G	A	A	G	G	G
On-target (EMX1_4)	C-->other bases	Untreated					0.04	0.06				0.15													
		ggX20					49.01	46.36				0.10													
		gX20					54.78	50.04				0.14													
		GX19					49.17	45.06				0.10													
		gX18					48.68	37.61				0.09													
		GX17					48.71	37.70				0.14													
			G	A	G	T	C	t	a	A	G	C	A	G	A	A	G	A	A	G	A	A	G	A	G
EMX1_1	C-->other bases	Untreated					0.04					0.05													
		ggX20					1.26					0.05													
		gX20					8.50					0.06													
		GX19					18.53					0.07													
		gX18					0.06					0.05													
		GX17					0.07					0.05													
			G	A	a	T	C	C	a	A	G	C	A	G	A	A	G	A	A	G	A	g	A	A	G
EMX1_2	C-->other bases	Untreated					0.08	0.08				0.07													
		ggX20					0.40	0.36				0.05													
		gX20					0.80	0.75				0.07													
		GX19					0.84	0.81				0.07													
		gX18					0.22	0.23				0.08													
		GX17					0.16	0.17				0.06													
			a	A	G	T	C	t	G	A	G	C	A	c	A	A	G	A	A	G	A	A	T	G	G
EMX1_3	C-->other bases	Untreated					0.02					0.07		0.06											
		ggX20					0.03					0.06		0.06											
		gX20					0.03					0.07		0.05											
		GX19					0.13					0.07		0.05											
		gX18					0.02					0.07		0.05											
		GX17					0.02					0.08		0.04											
			G	A	a	T	C	C	a	A	G	-	A	G	A	A	G	A	A	G	A	A	T	G	G
EMX1_5	C-->other bases	Untreated					0.11	0.05																	
		ggX20					1.09	1.11																	
		gX20					2.31	2.27																	
		GX19					0.96	0.96																	
		gX18					0.06	0.11																	
		GX17					0.08	0.12																	
			G	A	G	T	C	C	t	A	G	C	A	G	g	A	G	A	A	G	A	A	G	A	G
EMX1_6	C-->other bases	Untreated					0.02	0.04				0.04													
		ggX20					0.34	0.35				0.05													
		gX20					1.69	1.71				0.05													
		GX19					2.43	2.40				0.04													
		gX18					0.02	0.02				0.05													
		GX17					0.02	0.04				0.04													
			G	A	G	T	C	C	a	A	G	C	A	G	t	A	G	A	g	G	A	A	G	G	G
EMX1_7	C-->other bases	Untreated					0.03	0.06				0.06													
		ggX20					0.03	0.06				0.08													
		gX20					0.04	0.08				0.08													
		GX19					0.07	0.10				0.07													
		gX18					0.03	0.05				0.09													
		GX17					0.03	0.05				0.08													
			G	t	G	T	C	C	t	A	G	-	A	G	A	A	G	A	A	G	A	A	G	G	G
EMX1_8	C-->other bases	Untreated					0.05	0.03																	
		ggX20					0.10	0.09																	
		gX20					0.38	0.35																	
		GX19					0.64	0.57																	
		gX18					0.60	0.56																	
		GX17					0.62	0.60																	
			a	A	G	T	C	C	G	A	G	g	A	G	A	g	G	A	A	G	A	A	A	G	G
EMX1_9	C-->other bases	Untreated					0.05	0.16																	
		ggX20					0.04	0.14																	
		gX20					0.05	0.17																	
		GX19					0.06	0.18																	
		gX18					0.05	0.18																	
		GX17					0.05	0.16																	
			G	A	G	g	C	C	G	A	G	C	A	G	A	A	G	A	A	a	g	A	C	G	G
EMX1_10	C-->other bases	Untreated					0.14	0.10				0.13													
		ggX20					0.26	0.21				0.12													
		gX20					0.44	0.39				0.19													
		GX19					3.45	3.70				0.17													
		gX18					0.17	0.10				0.15													
		GX17					0.18	0.09				0.16													
			a	g	t	T	C	C	a	A	G	C	A	G	A	A	G	A	A	G	c	A	T	G	G
EMX1_11	C-->other bases	Untreated					0.06	0.05				0.07									0.06				
		ggX20					0.27	0.28				0.07									0.07				
		gX20					0.74	0.70				0.07									0.08				
		GX19					0.74	0.62				0.06									0.07				
		gX18					0.06	0.05				0.07									0.06				
		GX17					0.07	0.06				0.07									0.07				
			G	A	G	T	C	C	a	c	a	C	A	G	A	A	G	A	A	G	A	A	A	G	A
EMX1_12	C-->other bases	Untreated					0.08	0.26		0.11		0.11													
		ggX20					0.07	0.21		0.12		0.11													
		gX20					0.09	0.23		0.09		0.11													
		GX19					0.17	0.33		0.17		0.10													
		gX18					0.07	0.25		0.10		0.13													
		GX17					0.08	0.23		0.11		0.11													
			G	A	G	T	C	C	a	A	G	-	A	G	A	A	G	A	A	G	t	g	A	G	G
EMX1_13	C-->other bases	Untreated					0.08	0.12																	
		ggX20					0.06	0.11																	
		gX20					0.07	0.11																	
		GX19					0.08	0.13																	
		gX18					0.08	0.13																	
		GX17					0.07	0.13																	
			G	A	G	T	C	C	t	A	G	-	A	G	A	A	G	A	A	G	g	A	A	G	G
EMX1_14	C-->other bases	Untreated					0.06	0.13																	
		ggX20					0.07	0.19																	
		gX20					0.07	0.17																	
		GX19					0.05	0.13																	
		gX18					0.06	0.12																	
		GX17					0.05	0.14																	

(*Continued*)

EMX1_15	C-->other bases		G	A	a	T	C	C	a	A	G	C	A	G	g	A	G	A	A	G	A	A	G	G	A
		Untreated					0.04	0.07				0.05													
		ggX20					0.09	0.15				0.04													
		gX20					0.54	0.60				0.08													
		GX19					0.14	0.18				0.05													
		gX18					0.04	0.07				0.05													
		GX17					0.01	0.07				0.06													
EMX1_16	C-->other bases		G	t	a	c	C	a	G	A	G	-	A	G	A	A	G	A	A	G	A	g	A	G	G
		Untreated				0.06	0.06																		
		ggX20				0.05	0.05																		
		gX20				0.06	0.05																		
		GX19				0.05	0.05																		
		gX18				0.06	0.04																		
		GX17				0.06	0.05																		
EMX1_17	C-->other bases		G	A	G	T	C	C	c	A	G	C	A	a	A	A	G	A	A	G	A	A	A	A	G
		Untreated					0.10	0.19	0.09			0.07													
		ggX20					0.10	0.16	0.10			0.07													
		gX20					0.19	0.24	0.13			0.05													
		GX19					0.11	0.20	0.07			0.07													
		gX18					0.12	0.24	0.09			0.06													
		GX17					0.12	0.20	0.07			0.06													
EMX1_18	C-->other bases		a	A	G	T	C	C	a	A	G	t	-	G	A	A	G	A	A	G	A	A	A	G	G
		Untreated					0.06	0.09																	
		ggX20					0.05	0.09																	
		gX20					0.05	0.08																	
		GX19					0.09	0.11																	
		gX18					0.05	0.08																	
		GX17					0.05	0.09																	
EMX1_19	C-->other bases		a	A	G	T	C	C	a	t	G	C	A	G	A	A	G	A	g	G	A	A	G	G	G
		Untreated					0.03	0.07				0.10													
		ggX20					0.05	0.07				0.09													
		gX20					0.03	0.08				0.12													
		GX19					0.24	0.30				0.12													
		gX18					0.03	0.08				0.10													
		GX17					0.05	0.07				0.09													
EMX1_20	C-->other bases		G	A	G	T	C	C	t	A	G	-	A	G	A	A	G	A	A	a	A	A	G	G	G
		Untreated					0.05	0.12																	
		ggX20					0.21	0.26																	
		gX20					0.43	0.50																	
		GX19					0.50	0.57																	
		gX18					0.06	0.12																	
		GX17					0.05	0.12																	
EMX1_21	C-->other bases		G	A	G	T	C	C	c	t	-	C	A	G	g	A	G	A	A	G	A	A	A	G	G
		Untreated					0.16	0.08	0.07			0.03													
		ggX20					0.12	0.07	0.06			0.04													
		gX20					0.15	0.11	0.08			0.04													
		GX19					0.24	0.17	0.16			0.06													
		gX18					0.12	0.09	0.07			0.04													
		GX17					0.14	0.08	0.06			0.02													
EMX1_22	C-->other bases		a	c	G	T	C	t	G	A	G	C	A	G	A	A	G	A	A	G	A	A	T	G	G
		Untreated		0.14			0.04					0.11													
		ggX20		0.14			0.16					0.13													
		gX20		0.15			0.20					0.16													
		GX19		0.15			0.62					0.12													
		gX18		0.22			1.24					0.12													
		GX17		0.13			4.49					0.11													
EMX1_23	C-->other bases		G	A	G	T	t	C	c	A	G	a	A	G	A	A	G	A	A	G	A	A	G	A	G
		Untreated						0.06	0.08																
		ggX20						0.06	0.09																
		gX20						0.09	0.13																
		GX19						0.06	0.09																
		gX18						0.07	0.11																
		GX17						0.07	0.09																
EMX1_24	C-->other bases		G	A	G	T	C	C	t	A	a	-	A	G	A	A	G	A	A	G	c	A	G	G	G
		Untreated					0.05	0.18													0.11				
		ggX20					0.06	0.20													0.16				
		gX20					0.07	0.19													0.12				
		GX19					0.05	0.22													0.12				
		gX18					0.07	0.19													0.15				
		GX17					0.04	0.18													0.12				
EMX1_25	C-->other bases		c	A	G	T	C	C	a	A	a	C	A	G	A	A	G	A	g	G	A	A	T	G	G
		Untreated	0.11				0.05	0.11				0.11													
		ggX20	0.11				0.08	0.12				0.09													
		gX20	0.10				0.05	0.10				0.10													
		GX19	0.11				0.07	0.13				0.11													
		gX18	0.13				0.05	0.14				0.13													
		GX17	0.10				0.07	0.13				0.12													

RNF2

On-target (RNF2_1)	C-->other bases		G	T	C	A	T	C	T	T	A	G	T	C	A	T	T	A	C	C	T	G	A	G	G
		Untreated			0.06			0.07						0.06					0.03	0.07					
		ggX20			22.35			29.23						3.10					0.10	0.08					
		gX20			20.82			28.93						3.23					0.10	0.09					
		GX19			19.23			31.12						3.45					0.16	0.08					
		gX18			9.19			19.16						1.61					0.07	0.08					
		gX17			2.34			7.73						0.95					0.06	0.09					

(*Continued*)

FANCF

On-target (FANCF_2)	C-->other bases		G	G	A	A	T	C	C	C	T	T	C	T	G	C	A	G	C	A	C	C	T	G	G
		Untreated						0.06	0.10	0.04			0.03			0.13			0.13		0.05	0.04			
		ggX20						9.20	8.19	7.94			4.25			0.12			0.12		0.06	0.04			
		gX20						8.12	7.31	6.89			3.01			0.13			0.12		0.05	0.03			
		GX19						10.26	9.44	9.28			4.12			0.18			0.12		0.05	0.04			
		GX18						9.74	8.81	8.16			3.14			0.15			0.14		0.06	0.02			
		gX17						3.36	2.80	2.77			1.14			0.12			0.12		0.05	0.04			
			t	G	A	A	T	C	C	C	a	T	C	T	c	C	A	G	C	A	C	C	A	G	G
FANCF_1	C-->other bases	Untreated						0.07	0.10	0.08			0.03		0.04	0.04			0.07		0.04	0.07			
		ggX20						0.06	0.11	0.07			0.03		0.03	0.05			0.11		0.03	0.06			
		gX20						0.09	0.10	0.09			0.04		0.02	0.05			0.09		0.03	0.09			
		GX19						0.16	0.16	0.18			0.06		0.05	0.07			0.09		0.03	0.07			
		GX18						0.80	0.79	0.79			0.25		0.13	0.08			0.10		0.03	0.06			
		gX17						0.08	0.10	0.09			0.02		0.03	0.06			0.09		0.02	0.07			
			G	G	A	g	T	C	C	C	T	c	C	T	a	C	A	G	C	A	C	C	A	G	G
FANCF_3	C-->other bases	Untreated						0.06	0.09	0.05		0.08	0.03			0.06			0.07		0.06	0.14			
		ggX20						0.06	0.08	0.04		0.07	0.04			0.06			0.05		0.07	0.15			
		gX20						0.10	0.13	0.08		0.10	0.05			0.07			0.06		0.07	0.15			
		GX19						0.20	0.23	0.18		0.16	0.06			0.08			0.06		0.07	0.18			
		GX18						0.05	0.09	0.05		0.08	0.03			0.05			0.06		0.08	0.17			
		gX17						0.05	0.08	0.04		0.11	0.05			0.06			0.07		0.09	0.15			
			G	G	A	g	T	C	C	C	T	c	C	T	a	C	A	G	C	A	C	C	A	G	G
FANCF_4	C-->other bases	Untreated						0.06	0.05	0.05		0.05	0.06			0.06			0.03		0.03	0.06			
		ggX20						0.05	0.05	0.05		0.04	0.04			0.07			0.04		0.03	0.04			
		gX20						0.08	0.07	0.06		0.06	0.08			0.06			0.02		0.02	0.06			
		GX19						0.11	0.09	0.12		0.05	0.06			0.07			0.04		0.04	0.04			
		GX18						0.07	0.07	0.06		0.05	0.06			0.07			0.04		0.03	0.07			
		gX17						0.06	0.05	0.04		0.06	0.04			0.07			0.03		0.02	0.06			
			G	G	A	A	T	C	C	C	T	T	C	T	a	C	A	G	C	A	t	C	C	T	G
FANCF_5	C-->other bases	Untreated						0.09	0.07	0.05			0.03			0.07			0.03			0.03			
		ggX20						0.07	0.07	0.04			0.04			0.06			0.03			0.02			
		gX20						0.07	0.05	0.06			0.04			0.05			0.04			0.03			
		GX19						0.10	0.07	0.05			0.03			0.07			0.03			0.02			
		GX18						0.08	0.06	0.06			0.03			0.07			0.04			0.03			
		gX17						0.09	0.05	0.05			0.05			0.08			0.04			0.02			
			G	G	A	g	T	C	C	C	T	c	C	T	G	C	A	G	C	A	C	C	T	G	A
FANCF_6	C-->other bases	Untreated						0.04	0.04	0.04		0.02	0.04			0.09			0.06		0.02	0.04			
		ggX20						0.03	0.04	0.04		0.02	0.04			0.12			0.05		0.03	0.06			
		gX20						0.05	0.06	0.05		0.03	0.06			0.11			0.07		0.06	0.04			
		GX19						0.13	0.09	0.09		0.05	0.06			0.12			0.06		0.05	0.03			
		GX18						0.06	0.05	0.04		0.03	0.04			0.08			0.07		0.04	0.05			
		gX17						0.05	0.05	0.04		0.04	0.05			0.14			0.05		0.05	0.04			
			G	G	A	A	c	C	C	C	g	T	C	T	G	C	A	G	C	A	C	C	A	G	G
FANCF_7	C-->other bases	Untreated					0.03	0.07	0.07	0.06			0.03			0.20			0.05		0.03	0.07			
		ggX20					0.27	0.29	0.28	0.32			0.10			0.21			0.05		0.02	0.07			
		gX20					1.46	1.50	1.49	1.48			0.80			0.20			0.04		0.04	0.06			
		GX19					1.06	1.07	1.07	1.02			0.71			0.22			0.07		0.03	0.07			
		GX18					0.04	0.07	0.05	0.09			0.01			0.17			0.04		0.04	0.06			
		gX17					0.04	0.06	0.04	0.09			0.01			0.17			0.05		0.03	0.06			
			G	t	c	t	c	C	C	C	T	T	C	T	G	C	A	G	C	A	C	C	A	G	G
FANCF_8	C-->other bases	Untreated			0.02		0.03	0.04	0.02	0.05			0.03			0.11			0.03		0.02	0.03			
		ggX20			0.02		0.01	0.03	0.05	0.05			0.01			0.08			0.02		0.02	0.04			
		gX20			0.01		0.02	0.04	0.04	0.04			0.02			0.08			0.02		0.03	0.03			
		GX19			0.02		0.02	0.04	0.04	0.08			0.03			0.10			0.03		0.02	0.03			
		GX18			0.04		0.09	0.09	0.10	0.13			0.05			0.10			0.04		0.03	0.03			
		gX17			0.04		0.09	0.11	0.11	0.13			0.07			0.11			0.05		0.02	0.04			
			a	a	A	A	T	C	C	C	T	T	C	c	G	C	A	G	C	A	C	C	T	A	G
FANCF_9	C-->other bases	Untreated						0.07	0.02	0.04			0.05	0.06		0.05			0.04		0.05	0.06			
		ggX20						0.08	0.03	0.04			0.03	0.06		0.06			0.05		0.05	0.04			
		gX20						0.09	0.03	0.04			0.04	0.05		0.05			0.04		0.06	0.04			
		GX19						0.10	0.04	0.05			0.05	0.06		0.06			0.04		0.05	0.03			
		GX18						0.10	0.06	0.07			0.06	0.05		0.07			0.03		0.06	0.04			
		gX17						0.06	0.06	0.06			0.05	0.06		0.04			0.03		0.05	0.06			
			t	G	t	A	T	t	t	C	T	T	C	T	G	C	c	t	C	A	g	g	C	T	G
FANCF_10	C-->other bases	Untreated																							
		ggX20																							
		gX20																							
		GX19																							
		GX18																							
		gX17																							
			G	G	A	A	T	a	t	C	T	T	C	T	G	C	A	G	C	c	C	C	A	G	G
FANCF_11	C-->other bases	Untreated								0.03			0.05			0.22			0.03	0.05	0.05	0.10			
		ggX20								0.03			0.05			0.23			0.02	0.06	0.04	0.09			
		gX20								0.03			0.03			0.23			0.03	0.05	0.05	0.10			
		GX19								0.04			0.03			0.21			0.02	0.06	0.05	0.09			
		GX18								0.03			0.04			0.20			0.02	0.04	0.05	0.07			
		gX17								0.04			0.05			0.24			0.02	0.06	0.06	0.08			
			G	a	g	t	g	C	C	C	T	g	a	a	G	C	c	t	C	A	g	C	T	G	G
FANCF_12	C-->other bases	Untreated																							
		ggX20																							
		gX20																							
		GX19																							
		GX18																							
		gX17																							
			a	c	c	A	T	C	C	C	T	c	C	T	G	C	A	G	C	A	C	C	A	G	G
FANCF_13	C-->other bases	Untreated		0.07	0.06			0.04	0.04	0.04		0.06	0.05			0.10			0.03		0.08	0.04			
		ggX20		0.13	0.07			0.05	0.04	0.05		0.05	0.02			0.08			0.04		0.03	0.06			
		gX20		0.10	0.08			0.04	0.04	0.04		0.06	0.04			0.09			0.04		0.07	0.04			
		GX19		0.13	0.08			0.15	0.15	0.14		0.13	0.09			0.10			0.04		0.06	0.04			
		GX18		0.14	0.12			1.03	0.99	0.94		0.40	0.14			0.09			0.04		0.05	0.05			
		gX17		0.15	0.15			2.04	1.96	1.94		0.75	0.26			0.10			0.04		0.06	0.04			
			t	G	A	A	T	C	C	t	a	a	C	T	G	C	A	G	C	A	C	C	A	G	G
FANCF_14	C-->other bases	Untreated						0.09	0.05				0.04			0.09			0.06		0.08	0.06			
		ggX20						0.10	0.05				0.04			0.08			0.07		0.11	0.07			
		gX20						0.08	0.05				0.05			0.12			0.07		0.09	0.06			
		GX19						0.10	0.08				0.03			0.11			0.07		0.10	0.07			
		GX18						0.46	0.42				0.04			0.13			0.05		0.08	0.06			
		gX17						0.10	0.05				0.03			0.11			0.06		0.09	0.07			
			c	t	c	t	g	t	C	C	T	T	C	T	G	C	A	G	C	A	C	C	T	G	G
FANCF_15	C-->other bases	Untreated	0.03		0.04				0.05	0.02			0.02			0.06			0.02		0.01	0.03			
		ggX20	0.03		0.03				0.04	0.03			0.02			0.07			0.01		0.02	0.04			
		gX20	0.04		0.03				0.06	0.03			0.02			0.05			0.02		0.03	0.04			
		GX19	0.03		0.02				0.07	0.03			0.02			0.05			0.03		0.02	0.04			
		GX18	0.03		0.04				0.05	0.03			0.03			0.05			0.02		0.02	0.02			
		gX17	0.03		0.02				0.05	0.04			0.03			0.06			0.04		0.02	0.02			

(*Continued*)

HBB

On-target (HBB_1)	C-->other bases		C	T	T	G	C	C	C	C	A	C	A	G	G	G	C	A	G	T	A	A	C	G	G
		Untreated	0.05				0.08	0.03	0.05	0.04		0.04					0.08								
		ggX20	0.30				4.68	6.16	6.49	5.84		0.15					0.08								
		gX20	0.09				2.76	3.27	3.37	3.07		0.11					0.07								
		gX19	0.10				3.01	4.51	4.88	4.64		0.14					0.08								
		gX18	0.08				2.20	6.12	6.80	6.30		0.15					0.07								
		gX17	0.08				0.63	3.27	4.07	3.74		0.10					0.10								
HBB_2	C-->other bases		t	T	g	c	t	C	C	C	A	C	A	G	G	G	C	A	G	T	A	A	A	C	G
		Untreated				0.07		0.06	0.04	0.05		0.04					0.06								
		ggX20				0.06		0.08	0.04	0.06		0.04					0.09								
		gX20				0.08		0.09	0.07	0.08		0.03					0.05								
		gX19				0.42		0.89	0.84	0.86		0.07					0.06								
		gX18				0.07		0.12	0.10	0.11		0.05					0.06								
		gX17				0.07		0.08	0.05	0.06		0.05					0.08								
HBB_3	C-->other bases		g	c	T	G	C	C	C	C	A	C	A	G	G	G	C	A	G	c	A	A	A	G	G
		Untreated		0.07			0.06	0.06	0.11	0.03		0.07					0.14			0.09					
		ggX20		0.10			0.09	0.11	0.14	0.08		0.07					0.15			0.07					
		gX20		0.10			0.74	0.77	0.79	0.70		0.09					0.13			0.08					
		gX19		0.09			0.80	0.86	0.87	0.75		0.07					0.11			0.09					
		gX18		0.12			0.46	0.64	0.64	0.53		0.05					0.11			0.10					
		gX17		0.09			0.16	0.19	0.24	0.18		0.04					0.14			0.09					
HBB_4	C-->other bases		g	T	g	G	C	C	C	C	A	C	A	G	G	G	C	A	G	g	A	A	T	G	G
		Untreated					0.07	0.13	0.06	0.09		0.04					0.06								
		ggX20					0.10	0.11	0.06	0.10		0.05					0.04								
		gX20					0.08	0.12	0.07	0.09		0.04					0.06								
		gX19					0.14	0.20	0.13	0.16		0.07					0.08								
		gX18					0.10	0.24	0.17	0.20		0.08					0.06								
		gX17					0.84	1.61	1.58	1.53		0.16					0.05								
HBB_5	C-->other bases		a	T	T	G	C	C	C	C	A	C	g	G	G	G	C	A	G	T	g	A	C	G	G
		Untreated					0.12	0.19	0.73	0.40		0.16					0.20								
		ggX20					0.16	0.20	0.73	0.48		0.19					0.25								
		gX20					0.20	0.23	0.80	0.47		0.14					0.21								
		gX19					0.36	0.42	0.95	0.73		0.20					0.21								
		gX18					0.24	0.32	0.89	0.60		0.20					0.24								
		gX17					0.17	0.20	0.75	0.49		0.20					0.22								
HBB_6	C-->other bases		a	c	T	c	t	C	C	C	A	C	A	a	G	G	C	A	G	T	A	A	G	G	G
		Untreated		0.11		0.12		0.11	0.20	0.08		0.05					0.17								
		ggX20		0.09		0.14		0.09	0.24	0.09		0.05					0.19								
		gX20		0.12		0.13		0.13	0.23	0.14		0.04					0.22								
		gX19		0.10		0.14		0.13	0.22	0.09		0.05					0.17								
		gX18		0.12		0.15		0.14	0.26	0.11		0.06					0.22								
		gX17		0.10		0.16		0.11	0.24	0.10		0.04					0.19								
HBB_7	C-->other bases		t	c	a	G	C	C	C	C	A	C	A	G	G	G	C	A	G	T	A	A	G	G	G
		Untreated		0.03			0.07	0.07	0.09	0.05		0.05					0.08								
		ggX20		1.37			0.17	0.76	0.99	1.08		0.08					0.09								
		gX20		2.47			0.41	1.72	2.24	2.30		0.15					0.08								
		gX19		2.82			0.80	2.89	4.01	4.20		0.14					0.09								
		gX18		3.34			1.71	5.48	7.00	7.65		0.30					0.08								
		gX17		3.86			1.68	5.97	7.44	7.65		0.15					0.10								

HEK2

On-target (HEK2_2)	C-->other bases		G	A	A	C	A	C	A	A	A	G	C	A	T	A	G	A	C	T	G	C	G	G	G
		Untreated				0.05		0.05					0.03						0.03			0.19			
		ggX20				30.30		47.30					0.03						0.14			0.15			
		gX20				36.76		44.99					0.08						0.13			0.16			
		GX19				11.89		34.66					0.05						0.27			0.15			
		gX18				2.02		45.27					0.02						0.03			0.19			
		gX17				2.77		30.94					0.02						0.03			0.18			
HEK2_1	C-->other bases		G	A	A	C	A	C	A	A	t	G	C	A	T	A	G	A	t	T	G	C	C	G	G
		Untreated				0.11		0.09					0.09									0.16			
		ggX20				0.12		0.09					0.14									0.18			
		gX20				0.17		0.14					0.13									0.19			
		GX19				0.19		0.22					0.12									0.18			
		gX18				0.12		0.10					0.11									0.20			
		gX17				0.11		0.09					0.13									0.20			
HEK2_3	C-->other bases		a	A	c	t	c	C	A	A	A	G	C	A	T	A	t	A	C	T	G	C	T	G	G
		Untreated			0.07		0.09	0.37					0.24									0.24			
		ggX20			0.09		0.08	0.39					0.24									0.30			
		gX20			0.08		0.08	0.38					0.25									0.28			
		GX19			0.08		0.08	0.38					0.24									0.27			
		gX18			0.08		0.08	0.39					0.24									0.30			
		gX17			0.06		0.06	0.36					0.23									0.28			

HEK3

On-target (HEK3_2)	C-->other bases		G	G	C	C	C	A	G	A	C	T	G	A	G	C	A	C	G	T	G	A	T	G	G
		Untreated			0.15	0.47	0.39				0.15					0.08		0.06							
		ggX20			6.89	25.21	26.19				0.61					0.07		0.05							
		gX20			6.36	32.68	37.05				1.76					0.06		0.11							
		GX19			0.93	25.39	32.09				0.75					0.09		0.13							
		GX18			0.95	14.23	21.59				1.68					0.09		0.10							
		gX17			0.14	0.65	0.85				0.40					0.10		0.06							
HEK3_1	C-->other bases		a	G	C	t	C	A	G	A	C	T	G	A	G	C	A	a	G	T	G	A	G	G	G
		Untreated			0.13		0.04				0.06					0.12									
		ggX20			0.13		0.05				0.05					0.14									
		gX20			0.12		0.04				0.04					0.15									
		GX19			0.14		0.09				0.04					0.17									
		GX18			0.14		0.04				0.06					0.13									
		gX17			0.11		0.04				0.05					0.12									

(*Continued*)

HEK3_3	C-->other bases		G	t	g	g	C	c	c	A	g	a	G	A	G	C	A	C	G	T	G	t	G	G	G
		Untreated					0.08	0.04	0.08							0.09		0.12							
		ggX20					0.07	0.04	0.09							0.10		0.12							
		gX20					0.08	0.06	0.08							0.10		0.09							
		GX19					0.07	0.05	0.08							0.10		0.18							
		GX18					0.08	0.05	0.10							0.10		0.11							
		gX17					0.07	0.05	0.09							0.08		0.12							
HEK3_4	C-->other bases		c	a	C	C	C	A	G	A	C	T	G	A	G	C	A	C	G	T	G	c	T	G	G
		Untreated	0.07		0.06	0.08	0.04				0.02					0.13		0.05				0.04			
		ggX20	0.09		0.06	0.09	0.06				0.03					0.11		0.06				0.04			
		gX20	0.08		0.07	0.07	0.06				0.03					0.12		0.06				0.06			
		GX19	0.07		0.06	0.10	0.06				0.02					0.11		0.06				0.05			
		GX18	0.08		0.07	0.10	0.05				0.02					0.12		0.04				0.04			
		gX17	0.08		0.07	0.08	0.05				0.01					0.10		0.07				0.05			
HEK3_5	C-->other bases		c	G	g	C	C	c	a	A	C	T	G	A	G	C	A	a	G	T	G	A	T	G	G
		Untreated	0.17			0.09	0.14	0.08			0.06					0.17									
		ggX20	0.17			0.08	0.15	0.09			0.04					0.19									
		gX20	0.16			0.08	0.13	0.09			0.03					0.18									
		GX19	0.16			0.10	0.13	0.09			0.05					0.20									
		GX18	0.19			0.08	0.15	0.09			0.05					0.18									
		gX17	0.18			0.07	0.14	0.07			0.06					0.19									
HEK3_6	C-->other bases		a	G	a	C	C	A	G	A	C	T	G	A	G	C	A	a	G	a	G	A	G	G	G
		Untreated				0.09	0.09				0.06					0.17									
		ggX20				0.08	0.11				0.05					0.19									
		gX20				0.09	0.11				0.04					0.19									
		GX19				0.08	0.11				0.05					0.18									
		GX18				0.10	0.11				0.03					0.14									
		gX17				0.11	0.11				0.06					0.15									
HEK3_7	C-->other bases		G	G	C	C	a	c	t	c	a	T	G	g	c	C	A	C	a	T	a	c	T	G	G
		Untreated			0.38	0.18		0.06		0.18					0.30	0.30		0.07				0.06			
		ggX20			0.42	0.15		0.07		0.20					0.28	0.27		0.07				0.05			
		gX20			0.39	0.14		0.08		0.15					0.28	0.21		0.08				0.06			
		GX19			0.44	0.15		0.07		0.17					0.28	0.26		0.08				0.06			
		GX18			0.45	0.14		0.07		0.16					0.25	0.26		0.06				0.05			
		gX17			0.42	0.14		0.07		0.19					0.26	0.26		0.07				0.04			

HEK4

On-target (HEK4_1)	C-->other bases		G	G	C	A	C	T	G	C	G	G	C	T	G	G	A	G	G	T	G	G	G	G	G
		Untreated			0.17		0.08			0.23			0.07												
		ggX20			1.97		48.84			1.50			0.08												
		gX20			1.20		44.02			1.39			0.06												
		GX19			1.38		41.26			0.50			0.10												
		GX18			0.27		39.88			1.43			0.07												
		gX17			0.23		5.72			1.10			0.35												
HEK4_2	C-->other bases		G	G	C	A	C	T	G	C	t	G	C	T	G	G	g	G	G	T	G	G	T	G	G
		Untreated			0.14		0.04			0.11			0.91												
		ggX20			0.17		0.39			0.13			0.93												
		gX20			0.21		1.66			0.15			1.11												
		GX19			0.27		6.55			0.25			0.99												
		GX18			0.16		0.11			0.14			0.90												
		gX17			0.15		0.06			0.10			0.93												
HEK4_3	C-->other bases		G	G	C	A	C	T	G	C	a	-	C	T	G	G	A	G	G	T	t	G	T	G	G
		Untreated			0.09		0.05			0.07			0.05												
		ggX20			0.08		0.10			0.09			0.04												
		gX20			0.10		0.26			0.09			0.06												
		GX19			0.09		0.27			0.09			0.04												
		GX18			0.08		0.05			0.06			0.05												
		gX17			0.08		0.04			0.07			0.05												
HEK4_4	C-->other bases		G	G	C	t	C	T	G	C	G	G	C	T	G	G	A	G	G	g	G	G	T	G	G
		Untreated			0.05		0.05			0.29			0.14												
		ggX20			0.13		2.87			0.34			0.13												
		gX20			0.11		2.94			0.38			0.14												
		GX19			0.10		2.53			0.35			0.15												
		GX18			0.04		0.13			0.30			0.12												
		gX17			0.05		0.05			0.29			0.13												
HEK4_5	C-->other bases		a	G	C	A	C	T	G	C	a	G	a	T	G	G	A	G	G	a	G	G	C	G	G
		Untreated			0.09		0.03			0.11															
		ggX20			0.11		0.03			0.09															
		gX20			0.08		0.07			0.14															
		GX19			0.15		0.58			0.17															
		GX18			0.08		0.03			0.09															
		gX17			0.06		0.03			0.10															
HEK4_6	C-->other bases		G	G	C	A	C	T	G	C	G	G	C	a	G	G	g	a	G	g	a	G	G	G	G
		Untreated																							
		ggX20																							
		gX20																							
		GX19																							
		GX18																							
		gX17																							
HEK4_7	C-->other bases		t	G	C	A	C	T	G	C	G	G	C	c	G	G	A	G	G	a	G	G	T	G	G
		Untreated			0.24		0.10			0.38			0.14	0.08											
		ggX20			0.18		0.38			0.29			0.13	0.06											
		gX20			0.19		1.64			0.36			0.14	0.09											
		GX19			0.43		9.74			0.32			0.13	0.08											
		GX18			1.01		11.33			0.56			0.11	0.08											
		gX17			0.18		0.16			0.26			0.13	0.08											
HEK4_8	C-->other bases		G	G	C	A	C	T	-	g	G	G	C	T	G	a	A	G	G	T	a	G	A	G	G
		Untreated			0.08		0.03						0.09												
		ggX20			0.18		0.64						0.05												
		gX20			0.18		0.62						0.05												
		GX19			0.07		0.16						0.06												
		GX18			0.08		0.03						0.08												
		gX17			0.07		0.03						0.06												

(*Continued*)

HEK4_9	C-->other bases		G	G	C	A	C	T	G	t	G	G	C	T	G	c	A	G	G	T	G	G	A	G	G
		Untreated			0.11		0.03						0.04			0.03									
		ggX20			0.10		0.04						0.02			0.04									
		gX20			0.12		0.03						0.03			0.03									
		GX19			0.12		0.02						0.04			0.04									
		GX18			0.10		0.03						0.03			0.03									
		gX17			0.06		0.02						0.03			0.03									
HEK4_10	C-->other bases		t	G	C	t	C	T	G	C	G	G	C	a	G	G	A	G	G	a	G	G	A	G	G
		Untreated			0.08		0.18			0.07			0.07												
		ggX20			0.07		0.17			0.06			0.07												
		gX20			0.09		0.16			0.06			0.05												
		GX19			0.06		0.16			0.09			0.07												
		GX18			0.07		0.17			0.07			0.06												
		gX17			0.04		0.06			0.02			0.04												
HEK4_11	C-->other bases		a	G	C	A	C	T	G	C	a	G	C	T	G	G	g	a	G	T	G	G	A	G	G
		Untreated			0.16		0.05			0.15			0.08												
		ggX20			0.11		0.17			0.10			0.08												
		gX20			0.15		0.35			0.16			0.08												
		GX19			0.19		1.78			0.27			0.11												
		GX18			0.13		0.33			0.12			0.08												
		gX17			0.14		0.07			0.10			0.09												
HEK4_12	C-->other bases		G	G	C	A	C	T	G	a	G	G	g	T	G	G	A	G	G	T	G	G	G	G	G
		Untreated			0.07		0.04																		
		ggX20			0.27		1.09																		
		gX20			0.30		1.94																		
		GX19			0.07		1.09																		
		GX18			0.07		0.04																		
		gX17			0.10		0.03																		
HEK4_13	C-->other bases		G	G	C	A	C	T	G	g	G	G	C	T	G	G	A	G	a	c	G	G	G	G	G
		Untreated			0.12		0.13						0.12							0.21					
		ggX20			0.10		0.15						0.10							0.14					
		gX20			0.12		0.15						0.12							0.20					
		GX19			0.12		0.19						0.11							0.19					
		GX18			0.12		0.14						0.13							0.19					
		gX17			0.12		0.13						0.10							0.18					
HEK4_14	C-->other bases		a	G	g	A	C	T	G	C	G	G	C	T	G	G	g	G	G	T	G	G	T	G	G
		Untreated					0.05			0.29			0.03												
		ggX20					1.37			0.31			0.04												
		gX20					1.03			0.44			0.05												
		GX19					4.70			0.38			0.06												
		GX18					1.67			0.29			0.04												
		gX17					6.06			0.88			0.07												
HEK4_15	C-->other bases		G	G	C	A	C	T	G	C	a	a	C	T	G	G	A	a	G	T	G	a	T	G	G
		Untreated			0.11		0.06			0.11			0.02												
		ggX20			0.10		0.10			0.08			0.02												
		gX20			0.08		0.16			0.08			0.03												
		GX19			0.10		0.32			0.09			0.02												
		GX18			0.08		0.06			0.06			0.01												
		gX17			0.10		0.04			0.09			0.02												
HEK4_16	C-->other bases		G	G	C	A	C	T	G	g	G	G	t	T	G	G	A	G	G	T	G	G	G	G	G
		Untreated			0.16		0.18																		
		ggX20			0.69		2.90																		
		gX20			0.87		3.94																		
		GX19			0.29		3.17																		
		GX18			0.18		0.21																		
		gX17			0.15		0.15																		
HEK4_17	C-->other bases		G	c	C	A	C	T	G	C	a	G	C	T	a	G	A	G	G	T	G	G	A	G	G
		Untreated		0.11	0.05		0.05			0.16			0.04												
		ggX20		0.11	0.10		0.69			0.17			0.04												
		gX20		0.11	0.16		1.46			0.17			0.04												
		GX19		0.11	0.29		3.27			0.28			0.04												
		GX18		0.13	0.14		0.69			0.15			0.04												
		gX17		0.12	0.12		0.23			0.18			0.03												
HEK4_18	C-->other bases		G	c	C	A	C	T	G	C	G	a	C	T	G	G	A	G	G	a	G	G	G	G	G
		Untreated		0.16	0.06		0.06			61.49			0.05												
		ggX20		0.12	0.06		0.06			60.75			0.04												
		gX20		0.10	0.07		0.06			60.11			0.05												
		GX19		0.12	0.08		0.11			61.02			0.05												
		GX18		0.14	0.08		0.08			60.97			0.03												
		gX17		0.12	0.07		0.08			60.12			0.05												
HEK4_19	C-->other bases		G	G	C	A	C	T	G	-	G	G	C	T	G	G	A	G	G	c	G	G	G	G	G
		Untreated			0.03		0.06						0.05							0.08					
		ggX20			0.04		0.11						0.08							0.08					
		gX20			0.04		0.10						0.05							0.11					
		GX19			0.05		0.05						0.09							0.08					
		GX18			0.03		0.05						0.07							0.09					
		gX17			0.01		0.03						0.02							0.06					
HEK4_20	C-->other bases		a	G	C	t	C	T	G	C	G	G	C	a	G	G	A	G	t	T	G	G	A	G	G
		Untreated			0.22		0.03			0.22			0.10												
		ggX20			0.25		0.02			0.20			0.10												
		gX20			0.23		0.02			0.21			0.10												
		GX19			0.22		0.02			0.20			0.09												
		GX18			0.23		0.02			0.16			0.09												
		gX17			0.25		0.02			0.23			0.10												

5. Extended sgRNAs broaden the base editing window

a. Base-editing efficiencies of cytosine base editor with extended sgRNAs.

CBEs and ABEs are often limited by a narrow editing window of several nucleotide positions (typically, positions 12 to 17 (suboptimal) or positions 14 to 17 (optimal)) upstream of a protospacer adjacent motif (PAM) sequence in the single-stranded, non-target DNA strand exposed for the deaminase reaction at a target protospacer site. To adjust this editing window, CBEs, namely, Base Editor 3 (BE3) or Target-AID, have been combined with truncated sgRNAs in cultured human cells (Kim et al., 2017c) or with extended sgRNAs in E. coli (Banno et al., 2018), respectively. Here, I tested a series of extended sgRNAs to achieve base editing at positions upstream of the current editing window in human cells.

I transfected HEK293T cells with plasmids encoding BE3 and a conventional GX19 sgRNA (“G” or “g” indicates a matched or mismatched guanine, respectively, at the 5′ end, whereas “X19” indicates a 19-mer RNA sequence complementary with the protospacer DNA sequence) or eight extended sgRNAs with additional nucleotides in the 5′ terminus at 4 target sites (*HBB*, Site 18, Site 19, *HBB*-E2). To measure substitution frequency, I performed targeted deep sequencing at each target (Figure 21). GX19 sgRNAs, mostly, were highly active at position 14 to 17, but some extended sgRNAs (ggX20, gX21, gX22 and gX23) were active at positions 18 to 22, upstream of the canonical base editing window. I focused on the relative efficiencies at each base position of (the

substitution frequency at each position normalized to the maximum substitution frequency), averaged over two tested sites which have cytosine within canonical base editing window (Figure 22). I noticed several extended sgRNAs could expend the base editing window up to position 22 upstream of PAM.

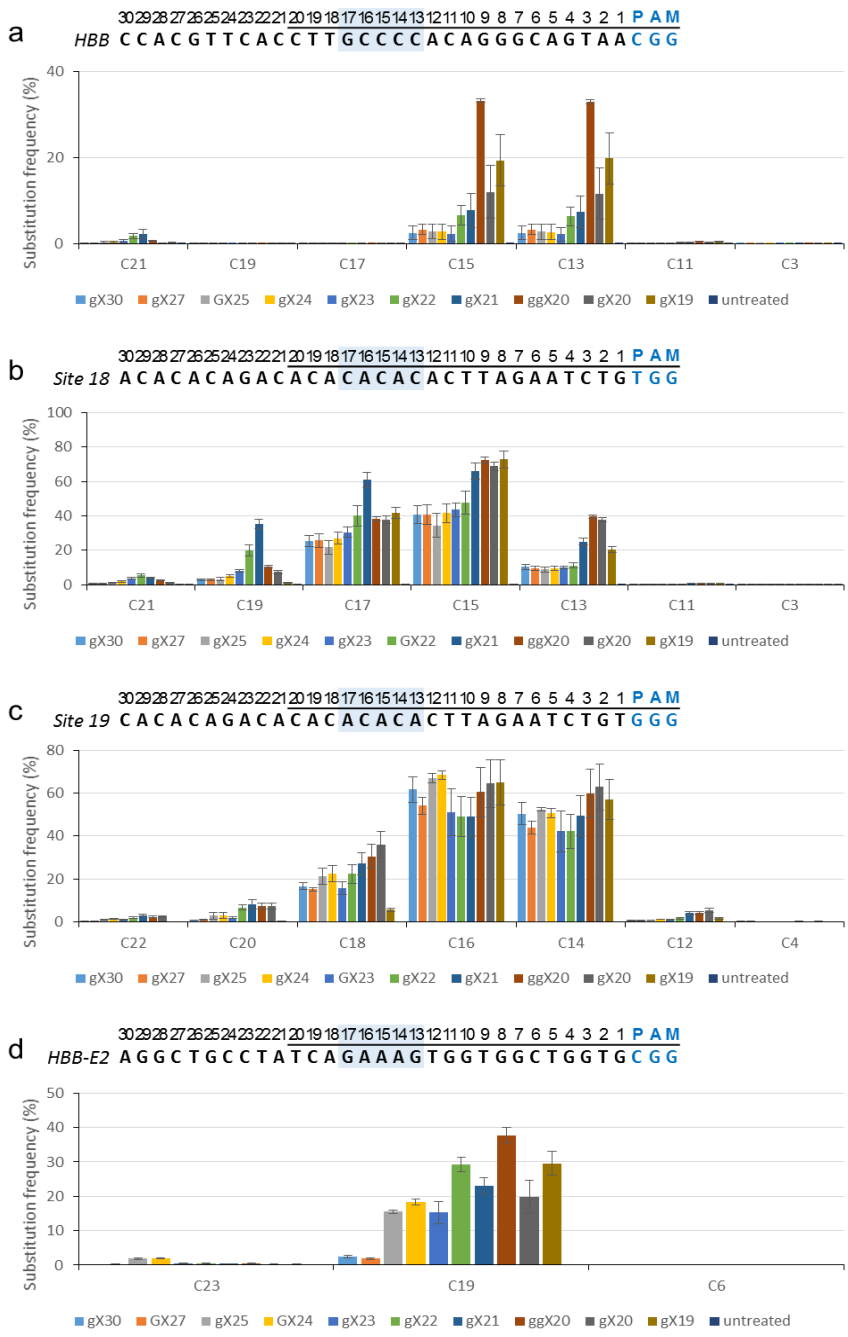


Figure 21. Base-editing efficiencies of cytosine base editor using extended sgRNAs. Extended sgRNAs with up to a 30-nucleotide spacer were tested in parallel with the conventional GX19 sgRNA at the *HBB* (a), Site 18 (b), Site 19 (c),

and *HBB*-E2 (d) target site in HEK293T cells. Substitution frequencies were measured at each base position by targeted deep sequencing.

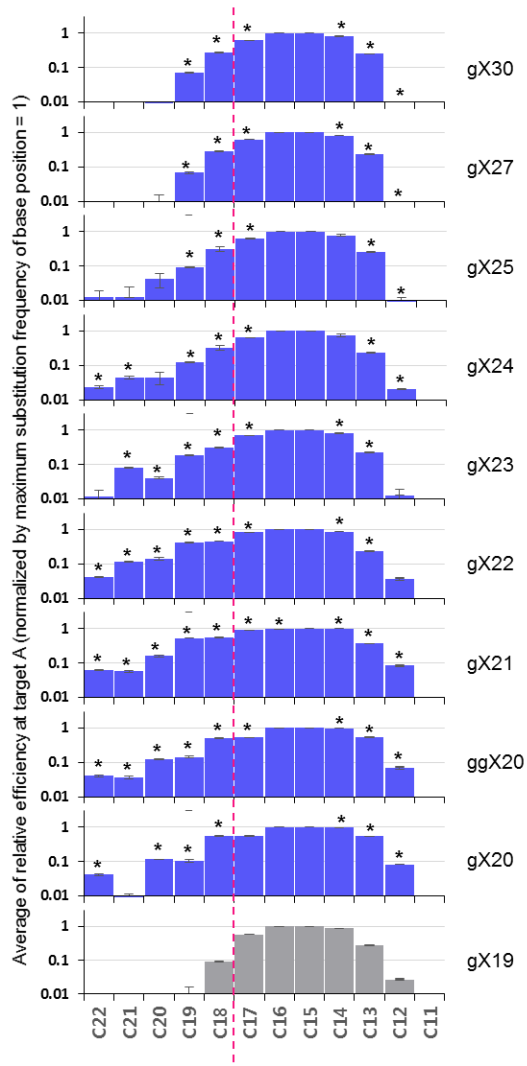


Figure 22. Relative efficiencies of BE3 at each base position. (The substitution frequency at each position normalized to the maximum substitution frequency) were averaged over two tested sites (Site 18 and Site 19). The red line indicates the boundary of the conventional activity window. Data are presented as mean± s.e.m. (n = 3 biologically independent samples).

b. Base-editing efficiencies of adenine base editor with extended sgRNAs.

To further characterization of base editing window using extended sgRNAs, I transfected HEK293T cells with plasmids encoding ABE7.10 and a conventional GX19 sgRNA targeted to the *HEK2* site used in previous studies (Gaudelli et al., 2017; Kim et al., 2017a) or eight extended sgRNAs with additional nucleotides at the 5′ terminus and performed targeted deep sequencing to measure substitution frequencies (Figure 23). As expected, the GX19 sgRNA was highly efficient with an editing frequency of 73% at position 16 but was only slightly active (<0.6%) at positions 18 and 19 upstream of the canonical base editing window. In contrast, extended sgRNAs with a few additional nucleotides at the 5′ terminus (GX20, gX21, gX22, and gX23) showed base editing activity at positions 18 and 19 with frequencies that ranged from 1.3% to 10%. sgRNAs with >23 base spacers were inactive at these positions and less active at positions even in the optimal editing window. Compared to the GX19 sgRNA, the GX20 sgRNA showed a 15-fold increase in the editing efficiency at positions outside of the canonical window (Figure 23b). Likewise, gX21, ggX20, gX22, and gX23 sgRNAs, targeted to four additional endogenous sites containing adenines at various positions, expanded the editing window, compared to gX19 and GX19 sgRNAs (Figure 24).

Taken together, these results show that sgRNAs with a few extra nucleotides, targeted to five endogenous sites (Table 7) in the human genome, induced A-to-G conversions at positions upstream of the current editing window,

demonstrating that it is feasible to expand the window of adenine base editing in human cells (Figure 25 and Table 8). These extended sgRNAs broaden the editing window by 25% or 50%, from positions 14–17 (four nucleotides) to positions 14–18 or 19 (five or six nucleotides).

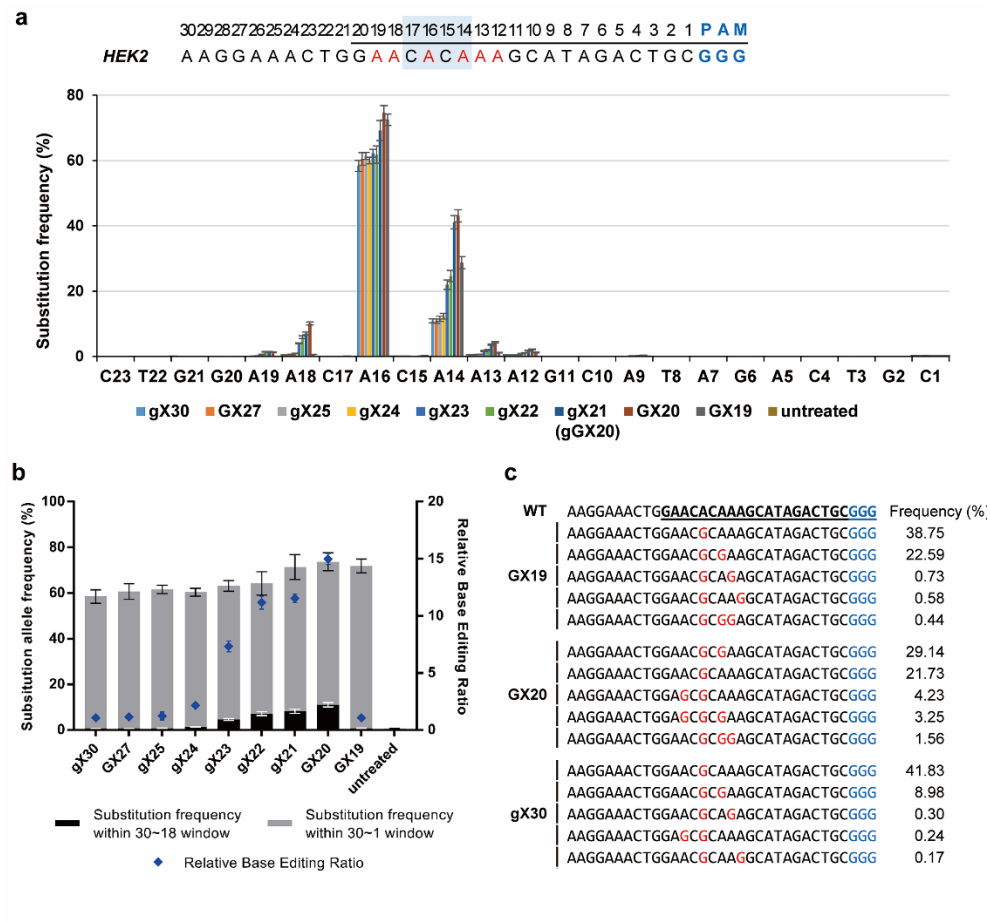


Figure 23. Base-editing efficiencies of adenine base editor using extended sgRNAs. (a) Extended sgRNAs with up to a 30-nucleotide spacer were tested in parallel with the conventional GX19 sgRNA at the *HEK2* target site in HEK293T cells. Substitution frequencies were measured at each base position by targeted deep sequencing. (b) The left y axis indicates the percentage of alleles that contain a base substitution at a target adenine within two windows (positions 1–30 (gray bars) or 18–30 (black bars)). The relative base editing ratio (right y axis), indicated by the blue dots, was calculated by dividing the substitution frequency within the 18–30 window by that within the 1–30 window. (c) Sequences of the most frequent

mutant alleles at the *HEK2* site. Substituted nucleotides are shown in red. The PAM is shown in blue.

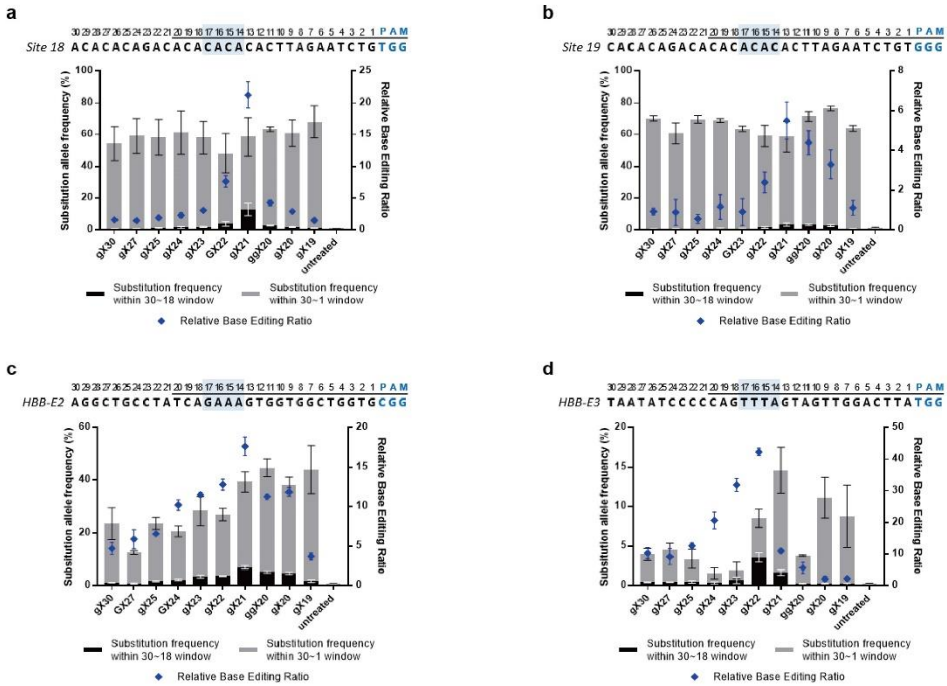


Figure 24. Activities of ABEs using extended sgRNAs in HEK293T cells. Base editing efficiencies of ABEs with extended sgRNAs at Site 18 (a), Site 19 (b), the *HBB*-E2 site (c), and the *HBB*-E3 site (d). (Left Y axis) Grey or black bars indicate the percentage of alleles that have a base-substitution mutation at the target adenine within two windows [positions 1-30 or 18-30 upstream from PAM], respectively. (Right Y axis) Blue dots indicate the relative base editing ratio, which was calculated by dividing the substitution allele frequency within the window of positions 18-30 by that within the window of positions 1-30. The PAM is shown in blue. Data are presented as mean ± s.e.m. (n = 3 biologically independent samples).

Table 7. ABE target sites in HEK293T cells.

No.	Sequence	Chromosome	Position	Direction	Reference
HEK2	GAACACAAAGCATAGACTGCGGG	chr5	87944780 - 87944799	+	3, 8
Site18	ACACACACACTTAGAATCTGTGG	chr1	184974900 - 184974919	+	3
Site19	CACACACACTTAGAATCTGTGGG	chr1	184974901 - 184974920	+	3
HBB-E2	TCAGAAAGTGGTGGCTGGTGTGG	chr11	5225630 - 5225649	-	
HBB-E3	CAGTTTAGTAGTTGGACTTAGGG	chr11	5225531 - 5225550	+	

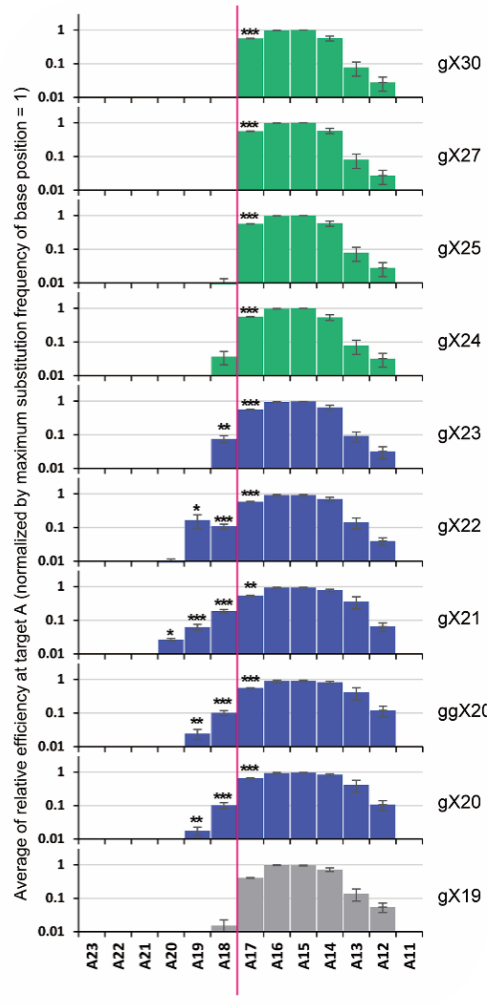


Figure 25. Relative efficiencies of ABE7.10 at each base position. (The substitution frequency at each position normalized to the maximum substitution frequency) were averaged over five tested sites (*HEK2*, Site 18, Site 19, *HBB*-E2, *HBB*-E3). The red line indicates the boundary of the conventional activity window. Data are presented as mean ± s.e.m. (n = 3 biologically independent samples). Two-tailed Student's t-test, *P < 0.05, **P < 0.01, ***P < 0.001.

Table 8. Relative editing efficiencies at each base position at 5 endogenous sites in the human genome. The substitution frequency at each position was normalized to the maximum substitution frequency.

	Position of target A												
sgRNA	A23	A22	A21	A20	A19	A18	A17	A16	A15	A14	A13	A12	A11
gX30	0.000	0.000	0.000	0.000	0.000	0.000	*** 0.561	0.981	1.000	0.570	0.077	0.028	0.000
gX27	0.000	0.000	0.000	0.000	0.000	0.000	*** 0.560	0.987	1.000	0.581	0.080	0.027	0.000
gX25	0.000	0.000	0.000	0.000	0.000	0.007	*** 0.571	0.980	1.000	0.585	0.079	0.028	0.000
gX24	0.000	0.000	0.000	0.000	0.000	0.037	*** 0.561	0.970	1.000	0.536	0.077	0.032	0.000
gX23	0.000	0.000	0.000	0.000	0.000	** 0.076	*** 0.565	0.947	0.970	0.649	0.092	0.032	0.000
gX22	0.000	0.000	0.000	0.006	* 0.165	*** 0.109	*** 0.584	0.915	0.922	0.692	0.142	0.040	0.000
gX21	0.000	0.000	0.000	* 0.027	*** 0.063	*** 0.187	** 0.540	0.934	0.940	0.774	0.358	0.065	0.000
ggX20	0.000	0.000	0.000	0.000	** 0.024	*** 0.101	*** 0.557	0.886	0.906	0.816	0.405	0.119	0.000
gX20	0.000	0.000	0.000	0.000	** 0.017	*** 0.101	*** 0.662	0.929	0.952	0.827	0.407	0.105	0.000
gX19	0.000	0.000	0.000	0.000	0.000	0.015	0.406	0.965	0.955	0.718	0.136	0.055	0.000

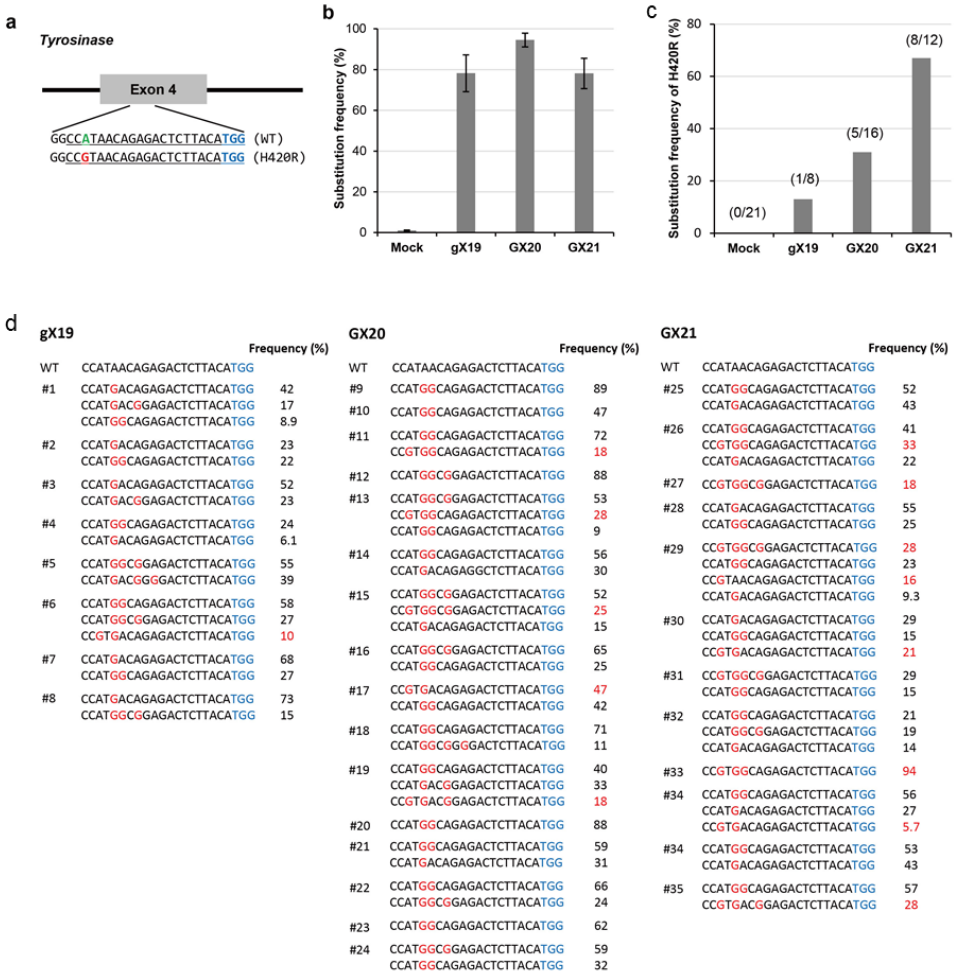
Two-tailed Student's t-test, *P < 0.05, **P < 0.01, ***P < 0.001.

6. Adenine base editing in mice using extended sgRNAs.

Based on the results obtained from human cell data, extended sgRNA was used for generating mouse models. Himalayan mutation (Ch/Ch) in the *Tyr* gene in mice is caused by a naturally occurring A-to-G single-nucleotide variation, resulting in a histidine-to-arginine amino acid change (H420R) in the tyrosinase protein, an essential enzyme for pigmentation in animals. The Himalayan mouse is characterized by a partial albinism in adults but is indistinguishable from albinos at birth (Green, 1961; Kwon et al., 1989). The target adenine in the *Tyr* gene is located at position 18, one nucleotide upstream of the canonical base editing window (Figure 26a), providing us with the right conditions for a proof-of-principle experiment in animals with extended sgRNAs. Each of three sgRNAs (gX19, GX20, and GX21) with ABE7.10 mRNA were delivered into mouse embryos by microinjection and genomic DNA isolated from the resulting blastocysts were analyzed. The three sgRNAs were almost equally efficient with average editing efficiencies of 78%, 95%, and 78%, respectively (Figure 26b). The Himalayan allele, however, was detected in just one out of eight embryos (13%) with a frequency of 10%, when the gX19 sgRNA was injected (Figure 26c,d). Thus, the overall frequency of the Himalayan allele obtained with the gX19 sgRNA was 1.3%. Notably, the mutant allele was created in 5 of 16 (31%) and 8 of 12 (67%) embryos with frequencies of up to 47% and 94%, respectively, when GX20 and GX21 sgRNAs, respectively, were used. As a result, the overall frequencies of the Himalayan allele obtained with these extended sgRNAs were 8.5% (GX20) and 20% (GX21) (Figure 27a).

I chose the GX21 sgRNA for further experiments and transplanted the resulting ABE-injected embryos into surrogate mothers to obtain several offspring carrying point mutations at the *Tyr* target site. Seven out of nine F0 mice (78%) harbored base substitutions, with frequencies that ranged from 12% to 100% (Figure 27b and Figure 28). One F0 mouse showed an albino phenotype in the eye even at birth (Figure 27c). Targeted amplicon sequencing revealed that this mouse was not genetically mosaic at the target site, carrying two Himalayan alleles. All of the other mutant mice were mosaic with at least three alleles or two non-Himalayan-mutant alleles. Altogether, base editing in 43 out of 45 (96%) embryos or pups with frequencies were ranged from 18% to 100% (Figure 26d and Figure 27d). I also verified the germline transmission of mutant alleles to F1 offspring (Figure 29).

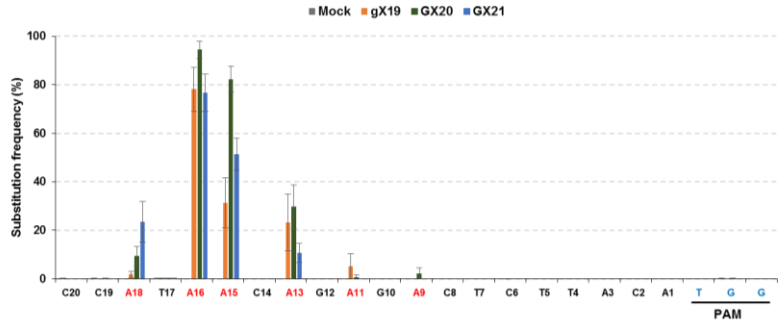
To assess off-target effects, I first used Cas-OFFinder (Bae et al., 2014) to identify potential off-target sites with up to 3-nucleotide mismatches in the mouse genome (Table 9). No off-target mutations were detectably induced at these sites in three *Tyr* mutant mice (Figure 30). Furthermore, to assess off-target effects in the *Tyr* mutant mouse (*Tyr* #4) carrying two Himalayan alleles, WGSs were carried. No off-target base editing was observed at potential off-target sites, identified 207,848 candidate sites that differed from the on-target site by up to seven mismatches, or identified 1,030,669 candidate sites that differed from the on-target site by up to five mismatches with an RNA or DNA bulge (Table 10). As a result, there were no off-target effects confirmed by deep-sequencing and WGS, revealing the high specificity of the *Tyr*-targeted ABE.



(By Seuk-Min Ryu & Kyoungmi Kim in Institute for Basic Science)

Figure 26. Adenine editing efficiencies in mouse embryos. (a) The target sequence at the *Tyrosinase* (*Tyr*) locus. The PAM sequence is shown in blue. The sgRNA target sequence is underlined. The targeted adenine in the wild-type sequence and the expected change in the sequence are shown in green and red, respectively. (b) Point mutation efficiencies associated with sgRNAs of different

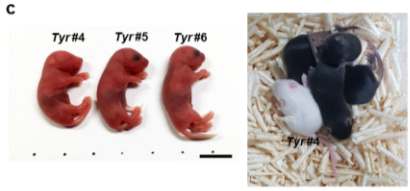
lengths (gX19, GX20, and GX21). Data are presented as mean ± s.e.m. (n = 8, 16 and 12 blastocysts, respectively). (c) Frequencies of blastocysts that carry the H420R mutation (Himalayan allele). (d) Alignments of mutant sequences from blastocysts after microinjection of ABE mRNA and *Tyr*-targeting sgRNAs of different lengths. The PAM and substitutions are shown in blue and red, respectively. The numbers in the column on the right indicate the frequency of each allele. WT, wild-type. H420R mutant allele frequencies are shown in red.



b

	420 421 422 423 424 425 426	Frequency (%)
	His Asn Arg Asp Ser Tyr Met	
WT	CCATAACAGAGACTCTTACATGG	
Tyr #4	CCGTGACGGAGACTCTTACATGG	49.0 (H420R, N421D, R422G)
	CCGTGGCAGAGACTCTTACATGG	48.9 (H420R, N421G)
Tyr #5	CCATGGCAGAGACTCTTACATGG	80.1 (N421G)
	CCATGACAGAGACTCTTACATGG	17.6 (N421D)
Tyr #6	CCATGGCAGAGACTCTTACATGG	44.6 (N421G)
	CCATGACAGAGACTCTTACATGG	27.9 (N421D)
	CCGTGGCAGAGACTCTTACATGG	26.6 (H420R, N421G)

c



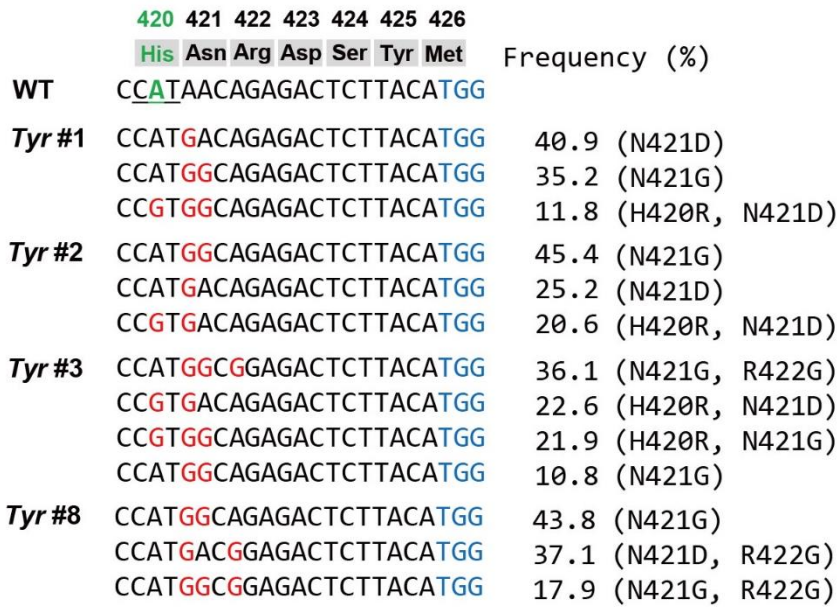
d

Target gene	No. of examined embryos	No. of two-cell stage embryos (%)	No. of blastocysts (%)	No. of transferred embryos	No. of Offspring (%)	Mutant ratio (%)	
						No. of mutants/total blastocysts or offspring (%)	No. of missense mutants/total blastocysts or offspring (%)
Buffer injection	41	39 (95)[a]	24 (62)[b]	NA	NA	0/21 (0)	0/21 (0)
Tyrosinase (gX19)	15	13 (87)[a]	8 (62)[b]	NA	NA	8/8 (100)	8/8 (100)
Tyrosinase (GX20)	24	21 (88)[a]	16 (76)[b]	NA	NA	16/16 (100)	16/16 (100)
Tyrosinase (GX21)	18	16 (89)[a]	12 (75)[b]	46	9 (20)[c]	19/21 (90)	19/21 (90)

NA: not applicable
a: Calculated from the number of examined embryos
b: Calculated from the number of developed 2-cell stage embryos
c: Calculated from the number of transferred embryos

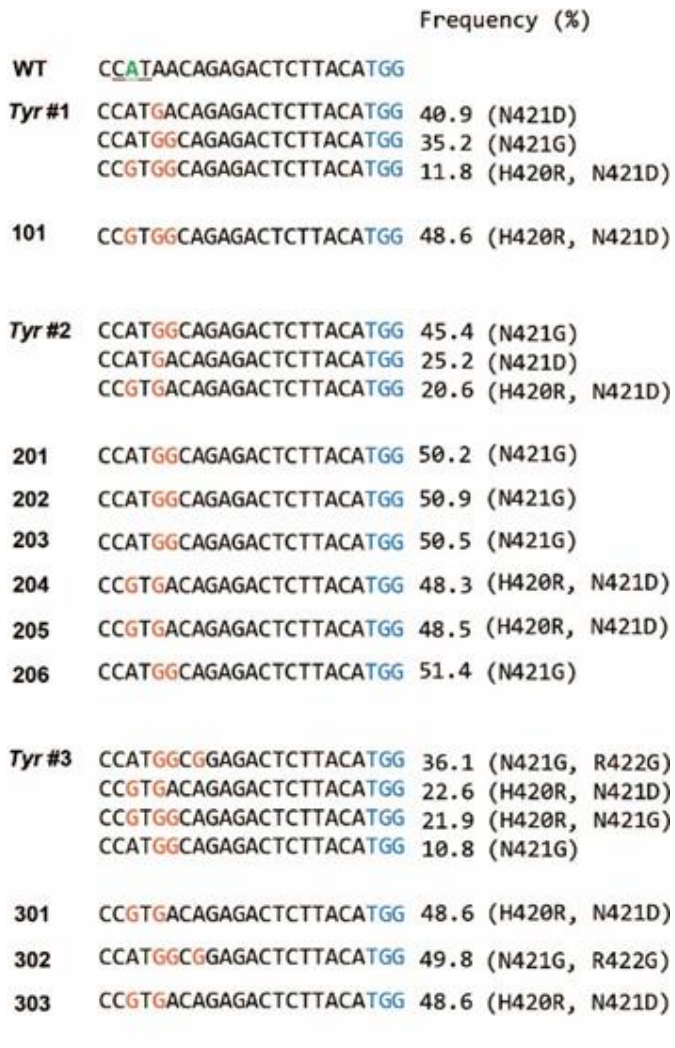
(By Seuk-Min Ryu & Kyoungmi Kim in Institute for Basic Science)

Figure 27. Frequencies of ABE-induced substitutions in the *Tyr* gene in mouse embryos and newborn pups. (a) Substitution frequencies at each position in the targeted locus. Results obtained with gX19, GX20, and GX21 sgRNAs are shown in yellow, green, and blue, respectively. Data are presented as mean ± s.e.m. (n = 8, 16, and 12 blastocysts, respectively). (b) Alignments of mutant sequences from newborn pups. The mutant sequences and the PAM site are shown in red and blue, respectively. The targeted position is shown in green. (c) *Tyr* mutant F0 pups that developed after microinjection of the ABE mRNA; one pup exhibited an albino phenotype in its eyes (*Tyr* #4, black arrow) and coat color. (d) Summary of adenine base editing results in mouse embryos or pups.



(By Seuk-Min Ryu & Kyoungmi Kim in Institute for Basic Science)

Figure 28. *Tyr* mutations in newborn pups. Alignments of mutant sequences from newborn pups. The altered nucleotides and the PAM site are shown in red and blue, respectively. The targeted position is shown in green.



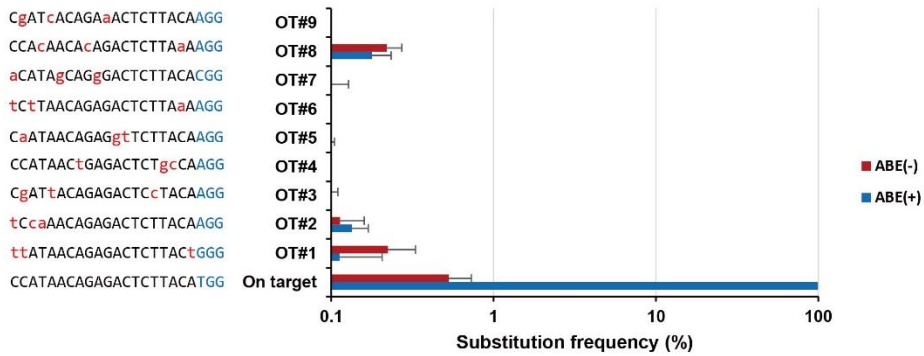
(By Seuk-Min Ryu & Kyoungmi Kim in Institute for Basic Science)

Figure 29. Germline transmission of *Tyr* mutant alleles. Germline transmission of mutant alleles to F1 pups (101, 201-206, and 301-303) from F0 *Tyr* mutant mice (*Tyr* #1, *Tyr* #2 and *Tyr* #3) was confirmed using targeted deep sequencing. The PAM site and altered nucleotides are shown in blue and red, respectively. The column on the right indicates frequencies of mutant alleles.

Table 9. Potential off-target sites in the mouse genome. Potential off-target sites were identified using Cas-OFFinder.

No.	Gene		Sequence	Chromosome	Position	Direction
On	*Tyr*	-	CCATAACAGAGACTCTTACATGG	chr7	87438023	-
OT1	Intergenic region		ttATAACAGAGACTCTTACtGGG	chr8	127915201	-
OT2	*Rsrc1*	Intron	tCcaAACAGAGACTCTTACAAGG	chr3	67007324	-
OT3	Intergenic region		CgATtACAGAGACTCcTACAAGG	chr5	79532301	+
OT4	*Adgrb3*	intron	CCATAACtGAGACTCTgcCAAGG	chr1	25463979	-
OT5	Intergenic region		CaATAACAGAGgtTCTTACAAGG	chr1	113881388	-
OT6	Intergenic region		tCtTAACAGAGACTCTTAaAAGG	chrX	65801977	+
OT7	Intergenic region		aCATAgCAGgGACTCTTACACGG	Chr18	52601483	-
OT8	*Ikzf1*	Intron	CCAcAACAcAGACTCTTAaAAGG	Chr11	11758527	-
OT9	Intergenic region		CgATcACAGAaACTCTTACAAGG	chr11	42283167	+

(By Seuk-Min Ryu & Kyoungmi Kim in Institute for Basic Science)



(By Seuk-Min Ryu & Kyoungmi Kim in Institute for Basic Science)

Figure 30. No off-target mutations at candidate sites in *Tyr* mutant mice. Potential off-target sites with up to 3 mismatches, relative to the wild-type sequence, were identified using Cas-OFFinder. Substitution frequencies at these potential off-target sites were measured using targeted deep sequencing. PAM sequences and mismatched nucleotides are shown in blue and red, respectively. Data are presented as mean ± s.e.m. (n = 9 animals).

Table 10. Whole genome sequencing to assess off-target effects in the *Tyr* mutant mouse. Genomes from the *Tyr* mutant mouse (*Tyr* #4) and a wild-type (WT) control mouse were sequenced using Illumina HiSeq X10. Unique single nucleotide variants (SNVs) in *Tyr* #4 by trimming out those in the WT control using the program 'Strelka' with the default 'eland' option. None of these SNVs other than the variations at the on-target site were found at potential off-target sites, identified using Cas-OFFinder, with up to 7 mismatches (207,848 sites) or with up to 5 mismatches and a DNA or RNA bulge (1,030,669 sites).

	Tyr #4
Number of all unique SNVs after Strelka analysis	15627
Number of passed unique SNVs after Strelka analysis	2547
Number of SNVs at potential off-target sites (207,848) with up to 7 mismatches	0
Number of SNVs at potential off-target sites (1,030,669) with up to 5 mismatches and a DNA or RNA bulge	0

(By Seuk-Min Ryu & Kyoungmi Kim in Institute for Basic Science)

Ⅳ. Discussion

The newly developed base editors are powerful tools for introducing point mutations at the desired locus. I tested various types of base editors, such as BE1, BE2, BE3, and ABE7.10 in PCR products, genomic DNA, mammalian cell lines, and mouse. Both *in vitro* and *in vivo* experiments were performed with high efficiency. Despite the successful base-editing, off-target effects were one major concern, especially for therapeutic use. Like all CRISPR-meditated genome-editing technologies, the base editors use binding activities of CRISPR-system, so they have the potential to operate on off-target DNA at the genomic locus. However, the genome-wide target specificities of CRISPR RNA-guided base editors remain unknown, owing to a lack of appropriate methods.

To develop a new method for assessing genome-wide specificities of CRISPR RNA-guided base editing, I modified Digenome-seq, *in vitro* Cas9-digested whole-genome sequencing. Genomic DNA was treated with the base editor and USER, uracil-specific excision reagents, *in vitro* to produce DNA double-strand breaks at uracil-containing sites. Off-target sites were then computationally identified from whole genome sequencing data. Testing seven different single guide RNAs (sgRNAs), I found that the rAPOBEC1–nCas9 base editor was highly specific, inducing cytosine-to-uracil conversions at only 18 ± 9 sites in the human genome for each sgRNA. Digenome-seq was sensitive enough to capture off-target sites with a substitution frequency of 0.1%. Notably, off-target sites of the base editors were often different from those of Cas9 alone, calling for

the independent assessment of their genome-wide specificities.

To minimize off-target effects, I changed the length of conventional sgRNAs which has 20nt complementarity to the target DNA. First, truncated sgRNAs, termed gX18 or gX17, were tested. Generally, substitutions events at off-target sites were decreased using truncated sgRNAs, but rather its efficiency increased in some off-target sites that have mismatches at 5' end. Second, I used sgRNAs with additional guanine bases at the 5'end, termed gX20 or ggX20. These extended sgRNAs discriminated on-target sites from off-target sites without sacrificing activities in on-target sites. Recently, several groups independently reported strategies for avoiding off-target effects, by use of Cas9 variants having higher specificity (Hu et al., 2018; Kleinstiver et al., 2016; Slaymaker et al., 2016), or delivery as ribonucleoprotein (Kim et al., 2014; Ramakrishna et al., 2014; Zuris et al., 2015) to minimize the operating time. Schemes to reduce the off-target of CRISPR nucleases were applicable to CRISPR deaminases, which is expected to be further improved regarding specificity.

CBEs and ABEs enabled to produce or eliminate single-nucleotide variations, a major source of pathogenic genetic diseases. Nevertheless, the activity window of base editors sometimes makes it impossible to change the bases at desired positions. Base editing windows of CBEs were known as positions 12 to 17 for BE1/2/3 (Komor et al., 2016) and positions 17 to 19 (optimal) or 13 to 19 (suboptimal) for Target-AIDs (Nishida et al., 2016), counting the position of a protospacer from the PAM. Base editing window of ABEs was positions 14 to 17

(optimal) or 12 to 17 (suboptimal) of upstream of PAM sequence at a target protospacer site (Gaudelli et al., 2017). I increased the length of sgRNA by extension the length of protospacer sequence up to 30nt, termed gX30~gX21, or with additional guanine bases at the 5' end, termed gX20 or ggX20.

BE3s using extended sgRNAs, like ggX20, gX21, gX22, and gX23, showed their activities at positions 13 to 22 from the PAM, beyond the base editing windows of Target-AIDs. Similarly, ABE7.10 with gX21, ggX20, gX22, and gX23 sgRNAs broadened the base editing window to positions 14 to 19. As *In vitro* experiments in HEK293 cells, ABE induced base editing *in vivo* in mouse at PAM-distal region compared to the conventional base editing window. Extended sgRNAs combined with engineered *S. pyogenes* Cas9 variants such as xCas9 and Cas9-NG having NG PAM (Hu et al., 2018; Nishimasu et al., 2018) or VQR-Cas9 having NGA PAM (Kim et al., 2017c), and either other types of CRISPR such as Cpf1 having TTTV PAM (Li et al., 2018), may further expand the targetable range of base editing.

In summary, results obtained using mismatched sgRNAs, Digenome-seq, and targeted deep sequencing, showed that BE3 deaminases were highly specific, catalyzing C-to-U conversions *in vitro* and base editing in human cells at a limited number of sites in the human genome. I found that BE3 and Cas9 off-target sites do not always coincide, justifying independent assessments of each tool. And I demonstrated that ABEs could be used to make disease models with single-nucleotide substitutions in mice. Also, the extended sgRNAs with a few additional

nucleotides at the 5′ end could expand the window of base editing *in vitro* and *in vivo*. I expect that these results and methods will accelerate the broad use of RNA-guided programmable deaminases in research and medicine.

Reference

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. Nat Commun *9*, 1911.

Bae, S., Park, J., and Kim, J.S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics *30*, 1473-1475.

Banno, S., Nishida, K., Arazoe, T., Mitsunobu, H., and Kondo, A. (2018). Deaminase-mediated multiplex genome editing in Escherichia coli. Nat Microbiol *3*, 423-429.

Bibikova, M., Beumer, K., Trautman, J.K., and Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. Science *300*, 764.

Bitinaite, J., Wah, D.A., Aggarwal, A.K., and Schildkraut, I. (1998). FokI dimerization is required for DNA cleavage. Proc Natl Acad Sci U S A *95*, 10570-10575.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science *326*, 1509-1512.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J., and Voytas, D.F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res *39*, e82.

Chapman, J.R., Taylor, M.R., and Boulton, S.J. (2012). Playing the end game: DNA double-strand break repair pathway choice. Mol Cell *47*, 497-510.

Cho, S.W., Kim, S., Kim, J.M., and Kim, J.S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol *31*, 230-232.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res *24*, 132-141.

Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., and Kuhn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol *33*, 543-548.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., *et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Crosetto, N., Mitra, A., Silva, M.J., Bienko, M., Dojer, N., Wang, Q., Karaca, E., Chiarle, R., Skrzypczak, M., Ginalski, K., *et al.* (2013). Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. Nat Methods *10*, 361-365.

Doyon, Y., McCammon, J.M., Miller, J.C., Faraji, F., Ngo, C., Katibah, G.E., Amora, R., Hocking, T.D., Zhang, L., Rebar, E.J., *et al.* (2008). Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. Nat Biotechnol *26*, 702-708.

Duan, J., Lu, G., Xie, Z., Lou, M., Luo, J., Guo, L., and Zhang, Y. (2014). Genome-wide identification of CRISPR/Cas9 off-targets in human genome. Cell Res

24, 1009-1012.

Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., Kii, E., and Alt, F.W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat Biotechnol *33*, 179-186.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol *31*, 822-826.

Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol *32*, 279-284.

Gabriel, R., Lombardo, A., Arens, A., Miller, J.C., Genovese, P., Kaeppel, C., Nowrouzi, A., Bartholomae, C.C., Wang, J., Friedman, G., *et al.* (2011). An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat Biotechnol *29*, 816-823.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67-71.

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. Nature *551*, 464-471.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., *et al.* (2014).

Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647-661.

Green, M.C. (1961). Himalayan - New Allele of Albino in Mouse. J Hered *52*, 73-&.

Guilinger, J.P., Thompson, D.B., and Liu, D.R. (2014). Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat Biotechnol *32*, 577-582.

Heigwer, F., Kerr, G., and Boutros, M. (2014). E-CRISP: fast CRISPR target site identification. Nat Methods *11*, 122-123.

Hess, G.T., Fresard, L., Han, K., Lee, C.H., Li, A., Cimprich, K.A., Montgomery, S.B., and Bassik, M.C. (2016). Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. Nat Methods *13*, 1036-1042.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., *et al.* (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol *31*, 827-832.

Hu, J.H., Miller, S.M., Geurts, M.H., Tang, W., Chen, L., Sun, N., Zeina, C.M., Gao, X., Rees, H.A., Lin, Z., *et al.* (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature *556*, 57-63.

Hur, J.K., Kim, K., Been, K.W., Baek, G., Ye, S., Hur, J.W., Ryu, S.M., Lee, Y.S., and Kim, J.S. (2016). Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins. Nat Biotechnol *34*, 807-808.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R., and Joung, J.K. (2013). Efficient genome editing in zebrafish

using a CRISPR-Cas system. Nat Biotechnol *31*, 227-229.

Jiang, F., Taylor, D.W., Chen, J.S., Kornfeld, J.E., Zhou, K., Thompson, A.J., Nogales, E., and Doudna, J.A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science *351*, 867-871.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol *31*, 233-239.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Kang, B.C., Yun, J.Y., Kim, S.T., Shin, Y., Ryu, J., Choi, M., Woo, J.W., and Kim, J.S. (2018). Precision genome engineering through adenine base editing in plants. Nat Plants *4*, 427-431.

Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H.R., Hwang, J., Kim, J.I., and Kim, J.S. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nat Methods *12*, 237-243, 231 p following 243.

Kim, D., Kim, J., Hur, J.K., Been, K.W., Yoon, S.H., and Kim, J.S. (2016a). Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nat Biotechnol *34*, 863-868.

Kim, D., Kim, S., Kim, S., Park, J., and Kim, J.S. (2016b). Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. Genome Res *26*, 406-415.

Kim, D., Lim, K., Kim, S.T., Yoon, S.H., Kim, K., Ryu, S.M., and Kim, J.S. (2017a). Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. Nat Biotechnol *35*, 475-480.

Kim, H., and Kim, J.S. (2014). A guide to genome engineering with programmable nucleases. Nat Rev Genet *15*, 321-334.

Kim, K., Ryu, S.M., Kim, S.T., Baek, G., Kim, D., Lim, K., Chung, E., Kim, S., and Kim, J.S. (2017b). Highly efficient RNA-guided base editing in mouse embryos. Nat Biotechnol *35*, 435-437.

Kim, S., Kim, D., Cho, S.W., Kim, J., and Kim, J.S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res *24*, 1012-1019.

Kim, Y., Kweon, J., Kim, A., Chon, J.K., Yoo, J.Y., Kim, H.J., Kim, S., Lee, C., Jeong, E., Chung, E., *et al.* (2013). A library of TAL effector nucleases spanning the human genome. Nat Biotechnol *31*, 251-258.

Kim, Y.B., Komor, A.C., Levy, J.M., Packer, M.S., Zhao, K.T., and Liu, D.R. (2017c). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat Biotechnol *35*, 371-376.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature *529*, 490-495.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-

stranded DNA cleavage. Nature *533*, 420-424.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., *et al.* (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583-588.

Kosicki, M., Tomberg, K., and Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. Nat Biotechnol *36*, 765-771.

Kuscu, C., Arslan, S., Singh, R., Thorpe, J., and Adli, M. (2014). Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nat Biotechnol *32*, 677-683.

Kwon, B.S., Halaban, R., and Chintamaneni, C. (1989). Molecular basis of mouse Himalayan mutation. Biochem Biophys Res Commun *161*, 252-260.

Li, X., Wang, Y., Liu, Y., Yang, B., Wang, X., Wei, J., Lu, Z., Zhang, Y., Wu, J., Huang, X., *et al.* (2018). Base editing with a Cpf1-cytidine deaminase fusion. Nat Biotechnol *36*, 324-327.

Liang, F., Han, M., Romanienko, P.J., and Jasin, M. (1998). Homology-directed repair is a major double-strand break repair pathway in mammalian cells. Proc Natl Acad Sci U S A *95*, 5172-5177.

Lin, Y., Cradick, T.J., Brown, M.T., Deshmukh, H., Ranjan, P., Sarode, N., Wile, B.M., Vertino, P.M., Stewart, F.J., and Bao, G. (2014). CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. Nucleic Acids Res *42*, 7473-7485.

Ma, H., Tu, L.C., Naseri, A., Chung, Y.C., Grunwald, D., Zhang, S., and Pederson, T. (2018). CRISPR-Sirius: RNA scaffolds for signal amplification in genome imaging. Nat Methods *15*, 928-931.

Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., and Chang, X. (2016). Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nat Methods *13*, 1029-1035.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. (2013). CRISPR RNA-guided activation of endogenous human genes. Nat Methods *10*, 977-979.

Maeder, M.L., Thibodeau-Beganny, S., Osiak, A., Wright, D.A., Anthony, R.M., Eichtinger, M., Jiang, T., Foley, J.E., Winfrey, R.J., Townsend, J.A., *et al.* (2008). Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell *31*, 294-301.

Makarova, K.S., Zhang, F., and Koonin, E.V. (2017a). SnapShot: Class 1 CRISPR-Cas Systems. Cell *168*, 946-946 e941.

Makarova, K.S., Zhang, F., and Koonin, E.V. (2017b). SnapShot: Class 2 CRISPR-Cas Systems. Cell *168*, 328-328 e321.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013a). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol *31*, 833-838.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013b). RNA-guided human genome engineering via

Cas9. Science *339*, 823-826.

Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. Nature *526*, 55-61.

Marraffini, L.A., and Sontheimer, E.J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature *463*, 568-571.

Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R., and Ploegh, H.L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat Biotechnol *33*, 538-542.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., *et al.* (2011). A TALE nuclease architecture for efficient genome editing. Nat Biotechnol *29*, 143-148.

Moscou, M.J., and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science *326*, 1501.

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y., *et al.* (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science *353*.

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O.O., Gootenberg, J.S., Mori, H., *et al.* (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science *361*, 1259-1262.

Park, J., Lim, K., Kim, J.S., and Bae, S. (2017). Cas-analyzer: an online tool for assessing genome editing results using NGS data. Bioinformatics *33*, 286-

288.

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., *et al.* (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods *10*, 973-976.

Ramakrishna, S., Kwaku Dad, A.B., Beloor, J., Gopalappa, R., Lee, S.K., and Kim, H. (2014). Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. Genome Res *24*, 1020-1027.

Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., *et al.* (2015). *In vivo* genome editing using Staphylococcus aureus Cas9. Nature *520*, 186-191.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., *et al.* (2013a). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380-1389.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013b). Genome engineering using the CRISPR-Cas9 system. Nat Protoc *8*, 2281-2308.

Rees, H.A., Komor, A.C., Yeh, W.H., Caetano-Lopes, J., Warman, M., Edge, A.S.B., and Liu, D.R. (2017). Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. Nat Commun *8*, 15790.

Rees, H.A., and Liu, D.R. (2018). Base editing: precision chemistry on the genome

and transcriptome of living cells. Nat Rev Genet.

Ryu, S.M., Koo, T., Kim, K., Lim, K., Baek, G., Kim, S.T., Kim, H.S., Kim, D.E., Lee, H., Chung, E., *et al.* (2018). Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. Nat Biotechnol *36*, 536-539.

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., *et al.* (2017). Diversity and evolution of class 2 CRISPR-Cas systems. Nat Rev Microbiol *15*, 169-182.

Shrivastav, M., De Haro, L.P., and Nickoloff, J.A. (2008). Regulation of DNA double-strand break repair pathway choice. Cell Res *18*, 134-147.

Singh, R., Kuscu, C., Quinlan, A., Qi, Y., and Adli, M. (2015). Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. Nucleic Acids Res *43*, e118.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science *351*, 84-88.

Sung, Y.H., Kim, J.M., Kim, H.T., Lee, J., Jeon, J., Jin, Y., Choi, J.H., Ban, Y.H., Ha, S.J., Kim, C.H., *et al.* (2014). Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. Genome Res *24*, 125-131.

Tsai, S.Q., Nguyen, N.T., Malagon-Lopez, J., Topkar, V.V., Aryee, M.J., and Joung, J.K. (2017). CIRCLE-seq: a highly sensitive *in vitro* screen for genome-wide CRISPR-Cas9 nuclease off-targets. Nat Methods *14*, 607-614.

Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J.A., Thapar, V., Reyon, D., Goodwin, M.J., Aryee, M.J., and Joung, J.K. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nat Biotechnol *32*, 569-576.

Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., *et al.* (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol *33*, 187-197.

Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D., and Holmes, M.C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature *435*, 646-651.

Wang, X., Wang, Y., Wu, X., Wang, J., Wang, Y., Qiu, Z., Chang, T., Huang, H., Lin, R.J., and Yee, J.K. (2015). Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. Nat Biotechnol *33*, 175-178.

Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., Chen, S., *et al.* (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat Biotechnol *32*, 670-676.

Yang, L., Briggs, A.W., Chew, W.L., Mali, P., Guell, M., Aach, J., Goodman, D.B., Cox, D., Kan, Y., Lesha, E., *et al.* (2016). Engineering and optimising deaminase fusions for genome editing. Nat Commun *7*, 13330.

Zalatan, J.G., Lee, M.E., Almeida, R., Gilbert, L.A., Whitehead, E.H., La Russa, M., Tsai, J.C., Weissman, J.S., Dueber, J.E., Qi, L.S., *et al.* (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell *160*, 339-350.

Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., Qiu, J.L., Wang, D., and Gao, C. (2017). Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. Nat Biotechnol *35*, 438-440.

Zuris, J.A., Thompson, D.B., Shu, Y., Guilinger, J.P., Bessen, J.L., Hu, J.H., Maeder, M.L., Joung, J.K., Chen, Z.Y., and Liu, D.R. (2015). Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing *in vitro* and *in vivo*. Nat Biotechnol *33*, 73-80.

국문초록

지난 몇 년 동안, CRISPR-Cas9 기술은 유전체 교정 연구뿐만 아니라 원하는 위치에 특정 기능을 도입할 수 있는 방법으로 개발되어 왔다. Cas9 유전자 가위에 의해 DNA 이중가닥 절단이 일어나고 유사한 염기 서열을 갖는 외부 DNA가 존재하면, 세포 내 DNA 이중가닥 절단 복구 메커니즘 중 하나인 상동 재조합 방법을 통해 점 돌연변이를 유도할 수 있다. 그러나 상동 재조합 기작과 경쟁적으로 작동하는 비상동성 말단 접합이 우세하기 때문에 세포 내에서 점 돌연변이를 유도할 수 있는 상동 재조합의 효율은 상대적으로 낮고, 세포 유형에 따라 상동 재조합의 효율이 다르게 나타난다.

이러한 제한점을 넘어 최근에 개발된 CRISPR 염기교정 유전자 가위는 염기 치환을 할 수 있는 새로운 방법을 제시하였다. 염기교정 유전자 가위는 DNA 이중가닥 절단 활성에 결함이 있는 *Sterptococcus pyogenes* Cas9 (dCas9 또는 nCas9) 과 단일 염기를 바꿀 수 있는 deaminase를 연결하여 만든 재조합 단백질이다. 염기교정 유전자 가위에는 시토신을 티민으로 치환 가능한 시토신 염기교정 유전자 가위와 아데닌을 구아닌으로 바꿀 수 있는 아데닌 염기교정 유전자 가위가 있다. 특정 위치에 이중가닥 절단을 일으켜 변이를 유도하는 유전자 가위들과는 다르게, 염기교정 유전자 가위는 DNA 이중가닥 절단 복구 메커니즘에 의존하지 않고 원하는 염기를 바꿀 수 있다. 이러한 큰 장점에도 불

구하고, 아직까지 염기교정 유전자 가위의 게놈 전반에 걸친 표적 특이성에 대해서는 밝혀진 바가 없었다.

이 연구에서는 보고된 DNA 이중가닥 절단이 일어난 곳을 확인할 수 있는 방법 중 하나인 Digenome-seq을 변형하여 시토신 염기교정 유전자 가위의 특이성을 확인하고자 하였다. 인위적으로 이중가닥 절단을 유도하기 위하여 DNA에 시토신 염기교정 유전자 가위와 함께 우라실을 특이적으로 자를 수 있는 효소들을 함께 처리하였다. 그 후, 게놈 전반에서 나타나는 시토신 염기교정 유전자 가위의 비표적 효과들을 확인하였고, 기존의 유전자 가위와 비교하여 시토신 염기교정 유전자 가위가 훨씬 정교하다는 것을 알 수 있었다. 관찰된 비표적 효과를 줄이기 위하여 가이드 RNA의 길이를 줄이거나 말단에 구아닌 염기를 추가하여 길이를 늘려보았고, 변형된 가이드 RNA를 이용하면 전반적으로 특이성이 향상됨을 확인하였다. 나아가 길이를 연장시킨 가이드 RNA를 사용한 시토신 또는 아데닌 염기교정 유전자 가위는 기존 염기 작동 범위보다 PAM으로부터 먼 부위에서도 작동할 수 있었다. 결과적으로 염기교정 유전자 가위의 변형된 가이드 RNA를 이용하면 염기교정 유전자가위의 정확성과 범용성을 개선할 수 있음을 입증하였다.

학 번: 2014-21246